A.A. Borovkov

ESTADÍSTICA MATEMÁTICA



Editorial Mir Moscú

A.A. Borovkov

ESTADÍSTICA matemática

Editorial Mir Moscú

A.A. Borovkov

ESTADÍSTICA *matemática*

Estimación de los parámetros
Verificación de las hipotesis
Capítulos adicionales



Editorial Mir Moscú

Traducido del ruso por A. Samojválov
Impreso en la URSS
На испанском языке

ISBN 5—03—000621—4.

Índice

Capítulo 4

Problemas estadísticos de dos muestras y más

Capítulo 5

Enfoque de los problemas de la estadística matemática desde el punto de vista de la teoría de los juegos

Prefacio

Este libro se basa en las conferencias de estadística matemática que el autor dictó durante muchos años en el tercer curso de la facultad de matemáticas de la Universidad de Novosibirsk. Con el andar del tiempo, el curso de conferencias ha sido varias veces modificado en busca de una variante que fuera, en la medida de lo posible, más armoniosa y accesible, y que al mismo tiempo correspondiera al estado moderno de esta ciencia. Se probaron distintas variantes, comenzando por un curso de carácter principalmente prescriptivo, con la exposición de los tipos básicos de problemas (construcción de estimaciones y criterios y estudio de sus propiedades), y terminando por un curso de carácter general, dedicado a la teoría de los juegos, en el que la teoría de las estimaciones y la verificación de las hipótesis eran no más que casos particulares de un enfoque único. A consecuencia del tiempo limitado (un semestre) no fue posible unificar dichas variantes íntimamente ligadas, cada una de las cuales poseía, por separado, defectos evidentes. En el primer caso, el conjunto de hechos concretos obstaculizaba el desarrollo de una opinión general en cuanto al objeto de estudio. La segunda variante carecía de resultados concretos sencillos y estaba sobrecargada de muchos conceptos nuevos, muy complejos, cuya asimilación constituía una tarea extraordinariamente difícil. Por lo visto, la más conveniente es la variante en la que la exposición de los elementos de la teoría de las estimaciones y de la teoría de verificación de las hipótesis concuerda con el mantenimiento consecutivo de la línea de búsqueda de los procedimientos óptimos.

Los capítulos fundamentales del libro se basan en el material unificado de las conferencias impartidas en tiempos diferentes y ampliadas a expensas de los apartados cuya presencia ha sido dictada por la propia lógica de exposición. El objetivo principal consiste en aclarar el estado actual de la

materia en concordancia con su accesibilidad máxima posible y la integridad y armonía matemática.

El libro comprende 5 capítulos y 8 suplementos.

En el capítulo 1 se estudian las propiedades (fundamentalmente asintóticas) de las distribuciones empíricas, que constituyen la base de la estadística matemática.

En los capítulos 2 y 3 se ofrecen, respectivamente, la teoría de las estimaciones y la teoría de verificación de las hipótesis estadísticas. Las primeras partes de cada uno de estos capítulos están dedicadas a la descripción de los posibles enfoques de la resolución de los problemas planteados, así como a la búsqueda de los procedimientos óptimos. Las segundas partes ofrecen la construcción de los procedimientos asintóticamente óptimos.

El capítulo 5 tiene esa misma estructura. En él se expone el enfoque general de los problemas de la estadística matemática desde el punto de vista de la teoría de los juegos.

El capítulo 4 está dedicado a los problemas relacionados con dos muestras y más.

Los suplementos del libro se hallan vinculados a las afirmaciones en el texto principal, cuya demostración sale fuera del marco de la exposición fundamental, ya por su carácter, ya por su dificultad.

El manual también contiene observaciones bibliográficas que no pretenden ser completas, pero que permiten seguir el surgimiento y el desarrollo de las principales tendencias de la estadística matemática. Además, por doquier donde ha sido posible, se ha dado preferencia a las alegaciones monográficas (como el tipo de literatura más accesible) y no a los artículos originales.

Hoy día existen bastantes manuales de estadística matemática. Entre ellos cabe destacar los cuatro siguientes, en cuyas páginas se expone un amplio material que refleja el estado actual de la materia: son los libros de H. Cramer [25], E. Lehmann [57], S. Zacks [95], I.A. Ibraguímov y R.Z. Jasminski [48]. Pero la máxima influencia en la escritura de la obra presente fue ejercida por las monografías [48] (algunas ideas de este libro se han utilizado en los §§ 23—25, 27—29 del cap. 2) y [57] (la exposición de los §§ 5—8 del capítulo 3 se asemeja, por su contenido, a los respectivos apartados de [57]). La demás exposición está poco relacionada, según su estructura, con los libros mencionados.

Hay muchas otras obras que ocupan un lugar notable en la literatura estadística (tales como los libros de Blackwell y Girshak [7], Kendall y Stuart [49, 50], Cox y Hinkly [23], Ferguson [33], Rao [76] y una serie de otros — no hay posibilidad de presentar su enumeración completa), pero por su espíritu y por la selección del material, estos trabajos se distinguen

considerablemente de la monografía que se ofrece a la atención de los lectores *).

A la par con los resultados y enfoques conocidos, en el libro presente se han incluido algunos apartados nuevos que simplifican la exposición del material, se han hecho varias mejoras metodológicas y se han utilizado algunos resultados nuevos, así como resultados que se publican por primera vez en la literatura monográfica.

A continuación se ofrece una descripción breve de la estructura metodológica del libro (véanse también el índice y los prefacios breves de cada uno de los capítulos).

En los §§ 1 y 2 del capítulo 1 se intrducen los conceptos de muestra y de distribución empírica y se establece el teorema de Glivenko — Cantelli, el cual puede considerarse como un hecho fundamental que constituye la base de las deducciones estadísticas.

En § 3 se introducen dos tipos de estadísticas (de los tipos I y II) que comprenden la inmensa mayoría de las estadísticas prácticamente interesantes, las cuales se definen como valores $G(\mathbf{P}_n^*)$ de las funcionales G (que satisfacen ciertas condiciones) de la distribución empírica $\mathbf{P}_n^*$. Más adelante, en los §§ 7 y 8 se establecen los teoremas del límite de distribución de dichas estadísticas. Esto simplifica la exposición posterior y permite no citar, para cada estadística concreta, prácticamente los mismos razonamientos que no se refieren, además, a la esencia de la cuestión.

En el § 5 han sido reunidos los teoremas auxiliares (que en el libro se denominan "teoremas de continuidad") sobre la convergencia de las distribuciones y la convergencia de sus momentos. Esto también simplifica la exposición posterior.

En el § 6 (no obligatorio en la primera lectura del libro) se establece que la función empírica de distribución $F_n^*(t)$ es un proceso poissoniano condicional, y se ofrece la enunciación del teorema (demostrado en el suplemento I) de la convergencia des proceso $\sqrt{n}(F_n^*(t) - F(t))$ hacia el puente browniano.

En el § 10 se introducen las distribuciones empíricas suavizadas que permiten aproximar no sólo la propia distribución, sino también su densidad.

En el § 3 del capítulo 2, dedicado a las estimaciones de los parámetros desconocidos, se introduce un método único de construcción de las estimaciones, denominado "método de sustitución". Este consiste en que la estimación θ^* para el parámetro θ, representado en forma de la funcional $\theta = G(\mathbf{P})$ de la distribución $\mathbf{P}$ de la muestra, es preciso buscarla en forma

---

*) En el año 1983 apareció un magnífico libro de *E. Lehmann* [58], en el cual, en adición a [57], se expone la actual teoría de estimación.

de $\theta^* = G(\mathbf{P}_n^*)$, donde $\mathbf{P}_n^*$ es la distribución empírica. Todas las estimaciones "razonables" usadas en la práctica son estimaciones de sustitución. La optimación de una estimación se alcanza eligiendo una funcional conveniente G. Si la estadística $\theta^* = G(\mathbf{P}_n^*)$ es de los tipos I ó II, los teoremas del capítulo 1 permiten establecer en seguida la validez de estas estimaciones y su normalidad asintótica. En los §§ 4 y 5, este enfoque es ilustrado por las estimaciones obtenidas mediante el método de momentos y el método de distancia mínima. Desde esas mismas posiciones también se podrían examinar las estimaciones de máxima verosimilitud (§ 6), pero su estudio inmediato da la posibilidad de obtener resultados más profundos, que serán necesarios ulteriormente.

La comparación de las estimaciones del capítulo 2 se realiza a base de dos enfoques: *estándar* o *medio cuadrático* (se comparan $\mathbf{M}_\theta (\theta^* - \theta)^2$ y *asintótico* (se comparan las varianzas de la distribución límite $\sqrt{n}(\theta^* - \theta)$ en la clase de estimaciones asintóticamente normales). En el caso paramétrico, esto permite destacar 3 tipos de estimaciones óptimas: estimaciones eficientes en las clases K_b, con un desplazamiento fijo b, y estimaciones bayesianas y minimax. A base de esos mismos principios se separan las clases de estimaciones *asintóticamente* óptimas en el enfoque asintótico. Para construir las estimaciones eficientes se utilizan los siguientes métodos tradicionales: el primero tiene carácter cualitativo y está vinculado al principio de suficiencia (§§ 12—14); el segundo se basa en las relaciones cuantitativas que se deducen de la desigualdad de Rao — Cramer (§ 16); y el tercero se halla relacionado con las consideraciones de invariación (§§ 17 y 19) que permiten reducir la clase de las estimaciones sometidas a examen.

Los §§ 20—30 están dedicados a la determinación de las estimaciones asintóticamente óptimas y al estudio de las propiedades asintóticas de la función de verosimilitud. El párrafo 20 contiene la desigualdad integral del tipo Rao — Cramer que permite, en particular, obtener criterios simples de carácter asintóticamente bayesiano y minimax de las estimaciones, así como fundamentar la separación de cierta subclase de estimaciones K_0 a la cual conviene limitarse en búsqueda de estimaciones asintóticamente eficientes. Esto da la posibilidad de establecer inmediatamente en el § 25, mediante el estudio de las propiedades asintóticas de las estimaciones de verosimilitud máxima, el carácter asintóticamente bayesiano y minimax de las estimaciones mencionadas, así como su eficiencia asintótica en K_0. Los párrafos 21—24 tienen carácter auxiliar. La estimación de los parámetros por intervalos se examina en los §§ 31 y 32 y también en el § 8 del capítulo 3.

El capítulo 3 está dedicado a la verificación de las hipótesis. En los §§ 1 y 2 se examina el caso de un número finito de hipótesis simples. Se

destacan (de un modo análogo a la teoría de estimación) tres tipos de criterios óptimos: los más potentes en sus subclases, los bayesianos y los minimax. Se establecen las relaciones entre estos criterios y se determina su forma evidente. Además, las consideraciones se basan en el principio bayesiano (y no en el lema de Neyman — Pearson) lo que, a nuestro juicio, simplifica la exposición y hace más comprensible el material. En el § 3 se examinan los enfoques asintóticos del cálculo de los criterios para verificar dos hipótesis simples y se realiza su comparación. En el § 4 se analiza el planteamiento general del problema sobre la verificación de dos hipótesis compuestas y se definen las clases de criterios óptimos (uniformemente más potentes, bayesianos y minimax). El párrafo 5 está dedicado a la búsqueda de criterios uniformemente más potentes en los casos cuando esto es posible. En los §§ 6 y 7 se resuelve el mismo problema, pero en las clases de criterios contraídos a base de consideraciones de no desplazamiento y de invariación. Además, al igual que en los §§ 1 y 2, las consideraciones se basan en el enfoque bayesiano. En el § 8 se construyen, con ayuda de los resultados obtenidos, los conjuntos confidenciales más exactos. En el § 9 se examinan los criterios bayesianos y minimax. Los párrafos 10 y 13 están dedicados al criterio de la relación de verosimilitud. Este criterio resulta uniformemente el más potente en muchos casos particulares y posee carácter asintóticamente bayesiano para conjeturas bastante amplias. El estudio de las propiedades de optimación asintótica del criterio de la relación de verosimilitud continúa en los §§ 15—17. En el § 11 se establece el valor óptimo de este criterio en los problemas del análisis sucesivo. Los párrafos 14 y 15 están dedicados a la búsqueda de criterios asintóticamente óptimos para verificar las hipótesis afines, y se ha encontrado su forma explícita simple para los principales problemas estadísticos.

Una particularidad importante de los tres primeros capítulos es el hecho de que en ellos se examinan tan sólo los problemas estadísticos relacionados con la utilización de una muestra.

Como ya fue señalado, el capítulo 4 del libro está dedicado a los problemas de dos muestras y más. A ellos pertenecen, antes que nada, los problemas sobre la homogeneidad (completa o parcial, §§ 1 y 2) y los problemas de regresión (§ 3) y del análisis de varianza (§ 4). A base de los resultados del capítulo 3, para los problemas de homogeneidad (en el caso paramétrico) se han construido los criterios asintóticamente óptimos, suponiendo que las hipótesis alternativas son semejantes a la hipótesis principal sobre la homogeneidad. Para los problemas de regresión (tanto para la regresión lineal como para la relacionada con las funciones arbitrarias) se han hallado, con ayuda de los resultados de los capítulos 2 y 3, las estimaciones eficientes de los parámetros desconocidos y se han construido los criterios para verificar las hipótesis principales. También han sido examinados los

llamados problemas de reconocimiento de imágenes (§ 5), los cuales, por lo visto, aparecen por primera vez en la literatura didáctica.

El capítulo 5 está dedicado al enfoque general de los problemas de estadística desde el punto de vista de la teoría de los juegos. Este enfoque contribuye a la formación de una opinión general acerca del objeto de estudio de la estadística matemática y permite generalizar muchos resultados de los capítulos 2 y 3. En el § 2 se exponen los conceptos y resultados principales de la teoría "ordinaria" de los juegos (se examinan únicamente los juegos de dos personas). En particular, se establecen las relaciones entre los tipos principales de estrategias óptimas: bayesianas, minimax y las uniformemente mejores en las subclases. En el § 3 se estudian los juegos estadísticos. En el § 4 se enuncia y se demuestra el llamado principio bayesiano que permite reducir el problema de búsqueda de la resolución estadística bayesiana a un problema mucho más fácil de construcción de la estrategia bayesiana para el juego ordinario de dos personas. En el § 5 se analizan los principios de suficiencia, de no desplazamiento y de invariación para construir las resoluciones uniformemente mejores en las subclases respectivas. Los párrafos 6—8 están dedicados a la búsqueda de las reglas decisivas asintóticamente óptimas. En el § 6 se estudian las estimaciones asintóticamente óptimas de los parámetros para la función arbitraria (y no sólo cuadrática) de pérdidas. En este caso se logra establecer los resultados semejantes a los del cap. 2 sobre la optimación asintótica de las estimaciones de verosimilitud máxima. En los § 7 y 8 se examinan los criterios asintóticamente óptimos para la función arbitraria de pérdidas. En el § 7 se demuestra el criterio asintóticamente bayesiano de la relacion de verosimilitud; en el § 8 se establece el indicio límite de optimación de los criterios para verificar las hipótesis semejantes (generalización de los resultados de los §§ 14 y 15 del cap. 3 para el caso de una función arbitraria de pérdidas).

Entre los Suplementos cabe destacar el Suplemento VIII donde se demuestran dos teoremas fundamentales de la teoría de los juegos estadísticos y cuya lectura exige una preparación matemática más alta.

El libro tiene muchas finalidades. Claro está que en su volumen completo, el mismo se asemeja más al programa mínimo para el curso de postgraduados de la especialidad de "Estadística Matemática", que a un libro de texto para los estudiantes. Pero en esta obra se prevé un sistema de medidas que facilitan su primera lectura y que la hacen accesible también para los estudiantes. Los párrafos de elevada dificultad o "más avanzados" en cuanto a su contenido están anotados con un asterisco y conviene omitirlos al leerlos por primera vez, así como el texto escrito con letra gallarda. Además, la exposición de los casos técnicamente más complicados, relaciona-

dos con el parámetro multidimensional, casi siempre se ofrece en apartados y párrafos independientes que también pueden ser omitidos.

Los profesores de los centros de enseñanza superior que ya conocen, al menos parcialmente, la asignatura pueden escoger del libro un conjunto de párrafos (puede haber muchas variantes) a base de los cuales (no es obligatorio utilizarlos por completo) es posible componer un curso semestral de estadística matemática. He aquí una de las variantes: §§ 1, 3 y 5 del capítulo 1; §§ 2—4, 6—12, 14, 16, (21, 23—25), 31 y 32 del capítulo 2; §§ 1, 2, 4, 5, 12 (13, 16) del capítulo 3. Los párrafos entre paréntesis están dedicados a los procedimientos asintóticamente óptimos. Según el grado de preparación de los estudiantes, es necessario organizar la enseñanza de dichos párrafos de la forma más accesible u omitirlos por completo.

La lectura del libro supone el conocimiento del curso de la teoría de las probabilidades conforme al volumen del manual de A.A. Borovkov [11]. Las remisiones a este libro, a diferencia de otras, aparecen en los lugares que el lector, por lo visto, debe conocer, y sirven fundamentalmente para hacer memoria.

La numeración de los párrafos en cada capítulo del libro es independiente, así como la de los teoremas (lemas, ejemplos, etc.) en cada párrafo. A fin de hacer más cómoda la lectura se utilizan diversos sistemas para las referencias a los teoremas, lemas, ejemplos, fórmulas, etc., según su alejamiento del pasaje que se lee. Si se hace una referencia al teorema 1 o a la fórmula (12) del párrafo que se lee, la misma se escribirá del siguiente modo: teorema 1, fórmula (12). Si se trata del teorema 1 y la fórmula (12) de uno de los párrafos precedentes de este capítulo (por ejemplo, del § 13), la referencia tendrá la forma siguiente: teorema 13.1, fórmula (13.12). Por último, si se hacen referencias a otro capítulo, aparecerá, además, el indicador del número de este último (primera cifra). Por ejemplo, el teorema 2.13.1 denota el teorema 1 del § 13 del capítulo 2, y la fórmula (2.13.12) denota la fórmula (12) del § 13 del capítulo 2. Eso mismo corresponde a la designación de los párrafos. La referencia al § 13 significa la remisión al § 13 de este capítulo, y la referencia al § 2.13 significa la remisión al § 13 del capítulo 2.

El signo $\triangleleft$ significa la terminación de la demostración.

Para facilitar la lectura del libro, al final de éste se da la lista de las principales designaciones y se expone el índice alfabético de materias.

A.A. Borovkov

Introducción

En el presente libro se exponen los fundamentos de la parte de las matemáticas que se llama *estadística matemática*. Para abreviar, esta última suele denominarse simplemente *estadística*. Sin embargo, conviene tener presente que tal abreviación sólo es posible cuando existe una buena comprensión mutua, puesto que, de por sí, el término "estadística" corresponde generalmente a un concepto algo distinto.

¿Qué representa la asignatura de estadística matemática? Se pueden citar diversas "definiciones" descriptivas que reflejan, en mayor o menor grado, el contenido de esta parte de las matemáticas. Una de las definiciones más simples y aproximadas se basa en la comparación relacionada con el concepto de selección de muestras de la población madre, así como con el problema de distribución hipergeométrica que se examina, por regla general, al principio del curso de teoría de las probabilidades. Conociendo la composición de la población madre, allí se estudian las distribuciones para la composición de una muestra aleatoria. Es un *problema directo* típico de la teoría de las probabilidades. No obstante, frecuentemente también es preciso resolver *problemas recíprocos* cuando se conoce la composición de la muestra y, basándose en ella, es necesario determinar cómo era la población madre. Tales tipos de problemas recíprocos son los que en realidad constituyen, hablando metafóricamente, la asignatura de estadística matemática.

Precisando algo esta comparación se puede decir lo siguiente: en la teoría de las probabilidades, conociendo la naturaleza de cierto fenómeno, aclaramos cómo se comportarán (cómo están distribuidas) unas u otras características sujetas a estudio, que pueden ser observadas en los experimentos. En la estadística matemática sucede al revés: como material de partida sirven los datos experimentales (generalmente las observaciones de las variables aleatorias) y es necesario adoptar uno u otro punto de vista

o tomar una decisión determinada sobre la naturaleza del fenómeno sujeto a examen. Ahora bien, aquí se trata de uno de los aspectos más importantes de la actividad humana: el proceso de conocimiento. La tesis de que "el criterio de la verdad es la práctica" está directamente relacionada con la estadística matemática, puesto que precisamente esta ciencia estudia los métodos (en el marco de los modelos matemáticos exactos) que permiten responder a la pregunta de si corresponde o no la práctica, representada en forma de los resultados del experimento, a la referida noción hipotética acerca de la naturaleza del fenómeno.

En este caso es necesario subrayar que, al igual que en la teoría de las probabilidades, nos interesarán no los experimentos que permiten sacar determinadas deducciones unívocas sobre los fenómenos examinados en la naturaleza, sino los experimentos cuyos resultados son sucesos aleatorios. Con el desarrollo de la ciencia, los problemas de tal género desempeñan un papel cada vez más importante, puesto que con el aumento de la precisión de los experimentos es cada vez más difícil evitar el "factor aleatorio" relacionado con diversos tipos de obstáculos y con nuestras limitadas posibilidades de medición y de cálculo.

La estadística matemática forma parte de la teoría de las probabilidades en el sentido de que cada problema de la estadística matemática es, en esencia, un problema (a veces muy peculiar) de la teoría de las probabilidades. Pero la estadística matemática, como tal, también ocupa una posición independiente en la clasificación de las ciencias. La estadística matemática puede considerarse como la ciencia del llamado comportamiento inductivo del hombre (y no sólo del hombre) en condiciones cuando éste, a base de su propia experiencia, debe tomar decisiones con las mínimas pérdidas para él *).

La estadística matemática también se llama teoría de las decisiones estadísticas, puesto que la misma puede ser caracterizada como la ciencia de las soluciones óptimas (las dos palabras siguientes requieren aclaración) basadas en los datos estadísticos (experimentales). Los planteamientos precisos de los problemas se darán posteriormente en el texto principal del libro. Aquí nos limitaremos a citar tres ejemplos de los problemas estadísticos más elementales y típicos.

Ejemplo 1. Para muchos artículos su plazo de servicio es uno de los parámetros principales que caracteriza la calidad. No obstante, el plazo de servicio de un artículo (digamos, de una bombilla eléctrica) es, por regla general, aleatorio y no se puede determinar de antemano. La experiencia muestra que si el proceso de producción es, en cierto sentido, homogéneo, los plazos de servicio ξ_1, ξ_2 ... de los respectivos artículos 1, 2 etc. pueden

*) Esta cuestión se examina más detalladamente en [46].

considerarse como magnitudes independientes igualmente distribuidas. El parámetro que nos interesa y que determina el plazo de servicio es natural identificarlo con el número $\theta = M\xi_1$. Uno de los problemas estándar consiste en determinar a qué es igual θ. Para hallar este valor se toman n artículos fabricados y los mismos se someten a comprobación. Sean $x_1, x_2, \ldots, x_n$ los plazos de servicio de dichos artículos comprobados. Sabemos que

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i \underset{c.s.}{\rightarrow} \theta$$

para $n \to \infty$. Por eso es natural esperar que, al ser n suficientemente grande, el número $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$ resultará próximo a θ y permitirá, en cierta medida, responder a las cuestiones planteadas. Es evidente que estamos interesados en que el número requerido de observaciones n sea el menor posible, y que nuestra estimación del número θ sea la más exacta posible (el aumento del parámetro θ, al igual que su reducción, conducirán a pérdidas materiales).

Ejemplo 2. Un radar explora, en los instantes de tiempo $t_1, t_2, \ldots, t_n$, una parte dada del espacio aéreo con el fin de localizar allí cierto objeto. Designemos por $x_1, \ldots, x_n$ los valores de las señales reflejadas que han sido recibidas por el radar. Si en la parte observada del espacio, el objeto que nos interesa no está presente, los valores de x_i pueden considerarse como variables aleatorias independientes distribuidas al igual que cierta variable aleatoria ξ cuya naturaleza está determinada por el carácter de las interferencias diferentes. Pero si en el transcurso de todo el período de observaciones, el objeto se encontraba en el campo de visión, entonces x_i contendrán, al igual que las interferencias, la señal "útil" a, y los valores de x_i se distribuirán como $\xi + a$. Ahora bien, si en el primer caso las observaciones de x_i tenían la función de distribución $F(x)$, en el segundo caso su función de distribución tendrá la forma $F(x - a)$. Por la muestra de $x_1, \ldots, x_n$ es preciso decidir cuál de estos dos casos tiene lugar, o sea, si existe o no, en la parte observada del espacio, el objeto que nos interesa.

En este problema será posible señalar, en cierto sentido, "la regla óptima decisiva" que resolverá el problema planteado, con errores mínimos. No obstante, el problema enunciado puede ser complicado del modo siguiente. Primero falta el objeto y luego, a partir de la observación de número θ desconocido, el mismo aparece. Hay que determinar, lo más exactamente posible, el instante θ de su aparición. Es el llamado "problema de desarreglo" que también tiene una serie completa de otras interpretaciones importantes para su aplicación.

Ejemplo 3. Cierto experimento se realiza al principio n_1 veces en condiciones A y luego n_2 veces en condiciones B. Designemos por $x_1, ..., x_{n_1}$ e $y_1, ..., y_{n_2}$ los resultados de estos experimentos en condiciones A y B, respectivamente. Es necesario contestar a la pregunta: ¿se reflejará el cambio de las condiciones del experimento en sus resultados? Con otras palabras, si designamos por $\mathbf{P}_A$ la distribución de x_i, $1 \leqslant i \leqslant n_1$, y por $\mathbf{P}_B$, la distribución y_i, $1 \leqslant i \leqslant n_2$, entonces la cuestión consistirá en contestar a la pregunta si se cumplirá o no la relación $\mathbf{P}_A = \mathbf{P}_B$.

Por ejemplo, si hay que determinar si influye o no cierto preparado en el desarrollo, digamos, de las plantas o los animales, entonces paralelamente se hacen dos series de experimentos (con el preparado y sin éste) cuyos resultados es preciso saber compararlos.

A menudo también surgen problemas más complejos cuando una cuestión análoga se plantea para muchas series de observaciones realizadas en condiciones diferentes. Si los resultados de tales observaciones dependen de las condiciones, suele ser necesario comprobar el distinto carácter de esta dependencia (el llamado problema de regresión).

El ejemplo 3 y los problemas más complejos anteriormente mencionados pertenecen a la clase de problemas estadísticos *con dos muestras y más*. Los mismos se examinan en el capítulo 4.

Podríamos continuar la lista de ejemplos de problemas estadísticos típicos, distintos en cuanto a su complejidad y a su esencia. No obstante, para ellos serán comunes las siguientes dos circunstancias:

1. No tendríamos ninguna dificultad si conociéramos las distribuciones de los resultados de las observaciones que figuran en los problemas.

2. En cada uno de estos problemas debemos, a base de los resultados de los experimentos, tomar cierta decisión en cuanto a la distribución de las observaciones disponibles (de aquí precisamente proviene la denominación "Teoría de las resoluciones estadísticas" mencionada más arriba).

En virtud de estas dos advertencias, para la exposición del material ulterior y, en particular, para la resolución de los problemos citados como ejemplos, adquiere importancia de principio el siguiente hecho. Según los resultados de las observaciones $x_1, ..., x_n$ de cierta variable aleatoria ξ, es posible, con grandes valores de n, restablecer, tan exactamente como se quiera, la distribución desconocida $\mathbf{P}$ de dicha variable aleatoria. La afirmación análoga también es válida para toda funcional $\theta = \theta(\mathbf{P})$ de esta distribución desconocida.

En este hecho se basa la estadística matemática. A él y a planteamientos más precisos de los problemas está dedicado el capítulo I.

CAPÍTULO I

Muestra. Distribución empírica. Propiedades asintóticas de las estadísticas.

En los §§ 1—4 se introducen los conceptos de muestra y de distribución empírica y se examinan sus propiedades elementales, principalmente asintóticas, que son la base de la estadística matemática.

En el § 5 se exponen los llamados teoremas de continuidad (sobre la convergencia de las distribuciones de las funciones de las sucesiones de variables aleatorias) que se utilizan en todo el libro.

Los §§ 6—10 están dedicados a propiedades asintóticas más finas de las distribuciones empíricas y al estudio de las distribuciones límites para los tipos principales de estadísticas.

§ 1. Concepto de muestra

El conjunto de resultados de las observaciones sirve de material inicial para toda investigación estadística. En los casos elementales, estos resultados no son más que los valores experimentales (obtenidos en las pruebas) de cierta variable aleatoria ξ. Ya hemos señalado que en los problemas de estadística, la distribución **P** de esta variable aleatoria se desconoce por lo menos parcialmente.

Supongamos que G es un experimento relacionado con la variable aleatoria ξ. Formalmente, para este experimento debemos construir un modelo matemático del cual forme parte el espacio probabilístico $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mathbf{P})$, y asignarle, de modo conveniente, la función medible que precisamente se denomina variable aleatoria ξ (véase [11]). El espacio $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mathbf{P})$, sin limitar la generalidad, puede considerarse "muestral" (véase [11]), o sea, podemos estimar que $\mathscr{X}$ es el espacio de los valores de $\xi(x) = x$. En este caso **P** se puede denominar distribución de ξ.

Si ξ es una variable aleatoria numérica, $\mathscr{X}$ es la recta numérica R; si ξ es un vector, $\mathscr{X} = R^m$, $m > 1$. En lo sucesivo tendremos en cuenta, por regla general, sólo estos dos casos, o sea, por $\mathscr{X}$ entenderemos R (caso uni-

dimensional) o bien R^m, $m > 1$ (caso multidimensional). En calidad de $\mathfrak{B}_{\mathscr{X}}$ se elige con más frecuencia el σ-álgebra de conjuntos de Borel [*].

Si se sabe de antemano que $\mathbf{P}$ está concentrada en la parte $B \in \mathfrak{B}_{\mathscr{X}}$ del espacio $\mathscr{X}$, por $\mathscr{X}$ puede resultar cómodo entender B, y por $\mathfrak{B}_{\mathscr{X}}$, la contracción del σ-álgebra $\mathfrak{B}_{\mathscr{X}}$ sobre B.

Examinemos n repeticiones independientes del experimento G (véase [11], p. 38) y designemos por $\mathbf{x}_1, \ldots, \mathbf{x}_n$ el conjunto de observaciones obtenidas. El vector

$$X_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

se llama *muestra de volumen n de la población con distribución* $\mathbf{P}$. A veces se utilizan variantes más breves o más completas de este término: "*muestra de la distribución* $\mathbf{P}$" o "*muestra simple de volumen n de la población madre con distribución* $\mathbf{P}$".

Simbólicamente, la relación "$\mathbf{X}_n$ es una muestra de la distribución $\mathbf{P}$" se escribirá, por medio del signo $\Subset$, del modo siguiente:

$$\mathbf{X}_n \Subset \mathbf{P}. \tag{1}$$

Tal forma de escritura también será utilizada para otras variables aleatorias. Por ejemplo, la relación

$$\xi \Subset \mathbf{P} \tag{2}$$

significará que ξ tiene la distribución $\mathbf{P}$. Tal uso del símbolo $\Subset$ se halla en correspondencia con (1), puesto que esta última ha sido determinada para cualquier n, en particular, para $n = 1$.

Si ξ y η son dos variables aleatorias (dadas, hablando en general, en diferentes espacios) con iguales distribuciones, designaremos este hecho por $\xi \underset{d}{=} \eta$, así que si X_n e Y_n son dos muestras de igual volumen de la distribución $\mathbf{P}$, podemos escribir $X_n \underset{d}{=} Y_n$.

En los segundos miembros de (1) y (2), en vez de la distribución $\mathbf{P}$ puede figurar, a veces, la función de distribución correspondiente a $\mathbf{P}$. Así que si $F(x) = \mathbf{P}((-\infty, x))$, la escritura de

$$X_n \Subset F$$

será idéntica a (1).

El propio concepto de "muestra de la población madre" también se

---

[*] Muchas partes del libro también serán válidas en una situación más general, cuando $\mathscr{X}$ es un espacio métrico arbitrario con un σ-álgebra $\mathfrak{B}_{\mathscr{X}}$ de conjuntos de Borel, o sea, con un σ-álgebra originada por los conjuntos abiertos de $\mathscr{X}$

encuentra al examinar modelos probabilísticos elementales relacionados con la extracción de bolas de una urna, en la definición clásica de la probabilidad (véase [11], § 2 cap. 1). Cabe señalar que la definición de la muestra, introducida más arriba, se halla en plena correspondencia con este concepto introducido anteriormente y, en esencia, coincide con él. Si x_i (o la variable aleatoria ξ) pueden adoptar sólo s valores $a_1, \ldots, a_s$, y las probabilidades de estos valores son racionales, o sea,

$$\mathbf{P}(\xi = a_j) = \frac{N_j}{N}, \quad \sum_{j=1}^{s} N_j = N,$$

entonces la muestra X_n puede representarse como el resultado del "muestreo con devolución" (en el sentido del cap. 1 [11]) de una urna con N bolas, entre las cuales N_1 bolas están marcadas con a_1, N_2 bolas con a_2, etc.

Como objeto matemático la muestra $X = X_n$ (el índice n será con frecuencia omitido) no es sino la variable aleatoria $(x_1, \ldots, x_n)$ con valores en el espacio "n-dimensional" $\mathscr{X}^n = \mathscr{X} \times \mathscr{X} \times \ldots \times \mathscr{X}$ y con una distribución que para $B = B_1 \times B_2 \times \ldots \times B_n$, $B_j \in \mathfrak{B}_{\mathscr{X}}$ se determina por las igualdades

$$\mathbf{P}(X \in B) = \mathbf{P}(x_1 \in B_1, \ldots, x_n \in B_n) = \prod_{i=1}^{n} \mathbf{P}(x_i \in B_i) \tag{3}$$

Con otras palabras, la distribución $\mathbf{P}$ sobre $\mathscr{X}$ es el producto directo múltiplo de n de las distribuciones "unidimensionales" dadas.

En lo que concierne a las designaciones de la distribución $\mathbf{P}$ y otras, nos sujetaremos a las siguientes acuerdos que ya hemos utilizado parcialmente en (3) y que nunca provocarán equivocaciones.

1. Utilizaremos el mismo símbolo (en particular, $\mathbf{P}$) para las distribuciones en $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}})$ y para el producto directo de estas distribuciones en $(\mathscr{X}^n, \mathfrak{B}^n_{\mathscr{X}})$ (véase (3)), donde $\mathfrak{B}^n_{\mathscr{X}}$ es el σ-álgebra de los conjuntos de Borel en $\mathscr{X}^n$. La diferencia será determinada tan sólo por el argumento de la función $\mathbf{P}$.

2. La probabilidad de llegada de la variable X, digamos, de $\mathfrak{B}^n_{\mathscr{X}}$ al conjunto B, a veces será cómodo designarla por $\mathbf{P}(B)$, y a veces por $\mathbf{P}(x \in B)$. Esto es lo mismo, ya que $\mathscr{X}^n$ es el espacio muestral de X.

3. Por último, utilizaremos el símbolo $\mathbf{P}$ para designar el concepto general de probabilidad (o sea, la probabilidad correspondiente a cualesquiera otras variables aleatorias sin concretizar el espacio probabilístico).

En virtud de (3) podemos considerar la muestra X como un suceso elemental en el espacio probabilístico muestral $(\mathscr{X}^n, \mathfrak{B}^n_{\mathscr{X}}, \mathbf{P})$ (véase [11] capítulo 3, § 2). Señalemos que en cuanto a la muestra X admitiremos una

interpretación doble de esta designación y del objeto: como variable aleatoria y como vector de los datos numéricos reales obtenidos en los experimentos realmente realizados. Como muestra la experiencia, tal interpretación doble es bien tolerable y no suscita equivocaciones, aunque admite la existencia simultánea de las notaciones que tienen la forma $\mathbf{P}(\mathbf{x}_1 < t) = F(t)$ y la forma $\mathbf{x}_1 = 0{,}74$, $\mathbf{x}_2 = 0{,}83$, etc.

La muestra es el objeto inicial principal en los problemas de la estadística matemática. Sin embargo, en la práctica, sus elementos $\mathbf{x}_1$, $\mathbf{x}_2$, ... no siempre, ni mucho menos, son independientes. En nuestros análisis tampoco excluiremos tal posibilidad. Además, para no hacer menciones adicionales, en caso de observaciones dependientes consideraremos que se trata de una muestra de volumen $n = 1$, mientras que las observaciones no son más que las coordenadas del vector $\mathbf{x}_1$ (en efecto, la naturaleza del espacio $\mathscr{X}$ es arbitraria).

En lo sucesivo tendremos que examinar a menudo las muestra $\mathbf{X}_n$ de volumen n indefinidamente creciente. En tales casos es cómodo suponer que se da la muestra $\mathbf{X}_\infty = (\mathbf{x}_1, \mathbf{x}_2 ...)$ de volumen infinito, y $\mathbf{X} = \mathbf{X}_n$ no es sino la población de sus primeras n coordenadas. Por muestra de volumen infinito $\mathbf{X}_\infty$ entenderemos el elemento del espacio probabilístico muestral $(\mathscr{X}^\infty, \mathfrak{B}_{\mathscr{X}}^\infty, \mathbf{P})$, donde $\mathscr{X}^\infty$ es el espacio de sucesiones $(x_1, x_2, ...)$; σ-álgebra $\mathfrak{B}_{\mathscr{X}}^\infty$ ha sido generada por los conjuntos $\bigcap_{i \leqslant N} \{x_i \in B_i\}$, $B_i \in \mathfrak{B}_{\mathscr{X}}$, $N = 1, 2, ...$; la distribución $\mathbf{P}$ posee la propiedad (3). Según el teorema de Kolmogórov ([11]), tal distribución siempre existe. Por consiguiente, la suposición sobre la existencia de la muestra $\mathbf{X}_\infty$ de volumen infinito de ningún modo limita la generalidad.

La propia sucesión infinita (muestra infinita) X_∞, en los estudios de carácter teórico-probabilístico puede interpretarse como un suceso elemental (compárese con [11]).

En los casos cuando necesitamos entender X_n como un subvector X_∞ escribiremos

$$X_n = [X_\infty]_n,$$

donde $[\cdot]_n$ es el operador de proyección de $\mathscr{X}^\infty$ en $\mathscr{X}^n$, determinado de modo evidente. Con arreglo a lo dicho anteriormente, la notación

$$X_\infty \in \mathbf{P}$$

significará que X_∞ es la muestra de volumen infinito de la distribución $\mathbf{P}$.

Si surge la necesidad de señalar especialmente el hecho de que no se trata de la distribución en $(\mathscr{X}^n, \mathfrak{B}_{\mathscr{X}}^n)$, sino en $(\mathscr{X}^\infty, \mathfrak{B}_{\mathscr{X}}^\infty)$ o en $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}})$ para $n < \infty$, también utilizaremos la designación $\mathbf{P}^\infty$ $(\mathbf{P}^n)$. La conservación de los índices superiores "∞" y "n" en todo el texto llevaría a designaciones muy complejas.

§ 2. Distribución empírica (caso unidimensional)

Sea dada la muestra $X = (x_1, ..., x_n) \in \mathbf{P}$, $x_i \in \mathscr{X} = R$. Examinemos la recta real R con σ-álgebra de los conjutos de Borel $\mathfrak{B}$ en la distribución discreta $\mathbf{P}_n^*$ sobre $(R, \mathfrak{B})$ concentrada en los puntos $x_1, ..., x_n$, para la cual la probabilidad del valor x_i se supone igual a $1/n$. En otros términos, para todo $B \in \mathfrak{B}$, según la definición,

$$\mathbf{P}_n^*(B) = \frac{\nu(B)}{n}, \qquad (1)$$

donde $\nu(B)$ es el número de elementos de la muestra X que se encuentran en el conjunto B. La distribución $\mathbf{P}_n^*$ se llama *distribución empírica* construida según la muestra X (o correspondiente a la muestra X). Esta distribución también puede representarse de la forma siguiente. Sea $\mathbf{I}_x(B)$ la distribución concentrada en el punto x:

$$\mathbf{I}_x(B) = \begin{cases} 1, & x \in B, \\ 0, & x \notin B; \end{cases}$$

entonces, evidentemente, $\nu(B) = \sum_{i=1}^{n} I_{x_i}(B)$,

$$\mathbf{P}_n^*(B) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_{x_i}(B). \qquad (2)$$

Está claro que para todo B de Borel, $\mathbf{P}_n^*(B)$ como función de la muestra es una variable aleatoria. Ahora bien, se trata de una función aleatoria de los conjuntos, o bien de una distribución aleatoria.

Supongamos ahora que $X_\infty \in \mathbf{P}$, $X_n = [X_\infty]_n$ y $n \to \infty$. Entonces obtendremos una sucesión de distribuciones empíricas $\mathbf{P}_n^*$. El hecho interesante consiste en que esta sucesión se aproxima indefinidamente a la distribución inicial $\mathbf{P}$ de la variable aleatoria sujeta a observación. Este hecho tiene importancia de principio para toda la exposición sucesiva, ya que el mismo muestra que la distribución desconocida $\mathbf{P}$ puede ser restablecida tan exactamente como se quiera, basándose en una muestra de volumen suficientemente grande.

Teorema 1. *Sea* $B \in \mathfrak{B}$ *y* $X_n = [X_\infty]_n \in \mathbf{P}$. *Entonces, para* $n \to \infty$

$$\mathbf{P}_n^*(B) \underset{c.s.}{\to} \mathbf{P}(B).$$

La convergencia con la probabilidad 1 aquí se sobreentiende con respecto a la distribución $\mathbf{P} = \mathbf{P}^\infty$ en $(R^\infty, \mathfrak{B}^\infty, \mathbf{P})$. Necesitamos la suposición $X_n = [X_\infty]_n$ para que las variables aleatorias $\mathbf{P}_n^*(B)$ se den en un solo espacio probabilístico.

Demostración. Examinemos la definición (2) y notemos que $I_{x_i}(B)$ son variables aleatorias independientes igualmente distribuidas, $\mathbf{M}I_{x_i}(B) = \mathbf{P}(I_{x_i}(B) = 1) = \mathbf{P}(x_i \in B) = \mathbf{P}(B)$. Como $\mathbf{P}_n^*(B)$ es la media aritmética de estas variables, nos queda hacer uso de la ley fuerte de los grandes números. ◁

El teorema 1 establece la convergencia de $\mathbf{P}_n^*(B)$ y $\mathbf{P}(B)$ en cada "punto" de B. No obstante, también tiene lugar una afirmación más fuerte de que tal convergencia es, en cierto sentido, uniforme respecto a B.

Designemos por $\mathfrak{F}$ la población de los conjuntos B que son semiintervalos de forma $[a, b)$ con extremos finitos o infinitos y volvamos a suponer que $X_n \Subset [X_\infty]_n$.

Teorema 2 (de Glivenko — Cantelli).

$$\sup_{B \in \mathfrak{F}} |\mathbf{P}_n^*(B) - \mathbf{P}(B)| \underset{c.s.}{\to} 0.$$

A decir verdad, con los nombres de Glivenko y Cantelli está relacionada una afirmación algo diferente, que se refiere a un concepto importante de la *función empírica de distribución*. Por definición, ésta es la función de distribución correspondiente a $\mathbf{P}_n^*$. En otros términos, se llama *función empírica de distribución* $F_n^*(x)$ la función

$$F_n^*(x) = \mathbf{P}_n^*((-\infty, x)).$$

La variable $nF_n^*(x)$ es igual al número de elementos de la muestra que son menores que x. En las condiciones reales, para construir $F_n^*(x)$ se utiliza a menudo el procedimiento siguiente. Los elementos de la muestra $(x_1, \ldots, x_n)$ se ordenan de manera creciente, o sea, de ella se forma la sucesión

$$x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$$

que se llama *serie variacional*. Entonces podemos suponer que

$$F_n^*(x) = \frac{k}{n} \text{ para } x \in (x_{(k)}, x_{(k+1)}),$$

donde k recorre los valores de 0 a n, $x_{(0)} = -\infty$, $x_{(n+1)} = \infty$. Evidentemente, $F_n^*(x)$ es una función escalonada que tiene saltos de $1/n$ en los puntos x_i si todos los valores de x_i son diferentes.

Sea $F(x) = \mathbf{P}(-\infty, x)$ la función de la distribución ξ (o x_1, que es lo mismo) y $X_n = [X_\infty]_n$. El teorema de Glivenko — Cantelli consiste en lo siguiente:

Teorema 2A. *Si* $n \to \infty$

$$\sup_x |F_n^*(x) - F(x)| \underset{c.s.}{\to} 0.$$

Más abajo omitiremos el índice n en las designaciones de F_n^* y escribiremos simplemente F^*.

Demostración del teorema 2A. Para abreviar supongamos primeramente que la función F es continua. Sea $\varepsilon > 0$ un número dado, arbitrariamente pequeño, de tal modo que el número $N = 1/\varepsilon$ sea entero. Cómo F es continua, podemos señalar los números $z_0 = -\infty, z_1, \ldots, z_{N-1}, z_N = \infty$ con los que

$$F(z_0) = 0,\ F(z_1) = \varepsilon = \frac{1}{N},\ \ldots,\ F(z_k) = k\varepsilon = \frac{k}{N},\ \ldots$$
$$\ldots,\ F(z_N) = 1.$$

Para $z \in [z_k, z_{k+1})$ son válidas las relaciones

$$F^*(z) - F(z) \leqslant F^*(z_{k+1}) - F(z_k) = F^*(z_{k+1}) - F(z_{k+1}) + \varepsilon, \tag{3}$$
$$F^*(z) - F(z) \geqslant F^*(z_k) - F(z_{k+1}) = F^*(z_k) - F(z_k) - \varepsilon.$$

Designemos por A_k el conjunto de sucesos elementales $\omega = X_\infty$ en los cuales $F^*(z_k) \to F(z_k)$. Según el teorema 1, $\mathbf{P}(A_k) = 1$. Por consiguiente, para cada $\omega \in A = \bigcap_{k=0}^{N} A_k$ se encontrará un valor de $n(\omega)$ tal, que para todos los valores de $n \geqslant n(\omega)$ se cumplirá

$$|F^*(z_k) - F(z_k)| < \varepsilon, \quad k = 0, 1, \ldots, N. \tag{4}$$

Pero junto con (3), dichas desigualdades contribuyen a que

$$\sup_z |F^*(z) - F(z)| < 2\varepsilon. \tag{5}$$

Así pues, esta relación tiene lugar para un valor arbitrario de $\varepsilon > 0$, para todos los valores de $\omega \in A$ y para todos los valores bastante grandes de $n \geqslant n(\omega)$. Como $\mathbf{P}(A) = 1$, el teorema para la función continua F se considera demostrado.

Para la función arbitraria $F(x)$, la demostración del teorema se realiza absolutamente igual. Se debe sólo hacer uso de la circunstancia siguiente: para toda $F(x)$ existe un número finito de puntos $-\infty = z_0 < z_1 < \ldots$ $\ldots < z_{N-1} < z_N = \infty$ con los que

$$F(z_{k+1}) - F(z_k + 0) \leqslant \varepsilon, \quad k = 0, 1, \ldots, N-1 \tag{6}$$

(para evidenciar podemos considerar que el conjunto $\{z_l\}$ contiene todos los puntos de los saltos de F que por sus valores superan, por ejemplo, $\varepsilon/2$). Absolutamente igual que en (3) obtenemos que para $z \in (z_k, z_{k+1}]$,

$$\begin{aligned} F^*(z) - F(z) &\leqslant F^*(z_{k+1}) - F(z_{k+1}) + \varepsilon, \\ F^*(z) - F(z) &\geqslant F^*(z_k + 0) - F(z_k + 0) - \varepsilon. \end{aligned} \tag{7}$$

A los conjuntos A_k, que se determinan como antes, les agregaremos los conjuntos A_k^+, $k = 0, 1, \ldots, N$ en los que $F^*(z_k + 0) \to F(z_k + 0)$. Entonces, según el teorema 1, $\mathbf{P}(A_k) = \mathbf{P}(A_k^+) = 1$, y en el conjunto $A = \bigcap_{k=0}^{N} A_k A_k^+$, $\mathbf{P}(A) = 1$, para valores de $n \geqslant n(\omega)$ bastante grandes será válida (4), así como las desigualdades

$$|F^*(z_k + 0) - F(z_k + 0)| < \varepsilon, \quad k = 0, 1, \ldots, N.$$

Junto con (7) estas desigualdades conducen a (5). ◁

El teorema 2A es un caso particular del teorema 2, ya que los conjuntos $(-\infty, x)$ pertenecen a $\mathfrak{F}$; por otro lado, el teorema 2 se obtiene fácilmente en calidad de corolario del teorema 2A, puesto que para $B = [a, b)$

$$|\mathbf{P}_n^*(B) - \mathbf{P}(B)| \leqslant |F_n^*(b) - F(b)| + |F_n^*(a) - F(a)|,$$

y, por consiguiente,

$$\sup_{B\in\mathfrak{F}} |\mathbf{P}_n^*(B) - \mathbf{P}(B)| \leqslant \sup_{a,\, b} [|F_n^*(b) - F(b)| + |F_n^*(a) - F(a)|] \underset{\text{c.s.}}{\to} 0.$$

Observación 1. Es fácil notar que los razonamientos de ese mismo género nos permiten, en calidad de población de los conjuntos $\mathfrak{F}$ en el teorema 2, tomar las poblaciones de los intervalos (a, b), de los segmentos $[a, b]$ y de sus uniones finitas (de número no mayor que cierto N).

Por otro lado, si en calidad de $\mathfrak{F}$ en el teorema 2 se toma una clase bastante rica de conjuntos, la afirmación del teorema deja de ser justa. Por ejemplo, si $\mathfrak{F}$ contiene las uniones de cualquier número finito de intervalos, entonces el conjunto $B_n = \bigcup_{k=1}^{n} (x_k - 1/n^2,\ x_k + 1/n^2) \in \mathfrak{F}$, $\mathbf{P}_n^*(B_n) = 1$ y para la distribución uniforme en $[0, 1]$, $\mathbf{P}(B_n) \leqslant 2/n$, así que

$$\sup_{B\in\mathfrak{F}} |\mathbf{P}_n^*(B) - \mathbf{P}(B)| \geqslant \mathbf{P}_n^*(B_n) - \mathbf{P}(B_n) \to 1.$$

Concluyendo este párrafo señalaremos que la representación (2) permite obtener para $\mathbf{P}_n^*$ teoremas sobre el comportamiento asintótico aún más exactos que los teoremas del tipo de Glivenko — Cantelli (estos resultados serán representados en los §§ 4 y 6). Para ilustrar las posibilidades que aquí existen recordemos que $\sum_{i=1}^{n} \mathbf{I}_{x_i}(B)$ en (2) es la suma de las variables aleatorias independientes e igualmente distribuidas en el esquema de Bernoulli

$$\mathbf{MI}_{x_i}(B) = \mathbf{P}(\mathbf{I}_{x_i}(B) = 1) = \mathbf{P}(B),$$
$$\mathbf{MI}_{x_i}^2(B) = \mathbf{P}(B), \quad \mathbf{DI}_{x_i}(B) = \mathbf{P}(B)(1 - \mathbf{P}(B)).$$

Por eso, del teorema central del límite se deduce inmediatamente la afirmación siguiente:

Teorema 3. $\mathbf{P}_n^*(B)$ *es representable en la forma*

$$\mathbf{P}_n^*(B) = \mathbf{P}(B) + \frac{\zeta_n(B)}{\sqrt{n}}, \tag{8}$$

donde la distribución $\zeta_n(B) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{I}_{x_i}(B) - \mathbf{P}(B))$ *converge hacia la distribución normal con los parámetros* $(0, \mathbf{P}(B)(1 - \mathbf{P}(B))$.

El estudio ulterior de $\mathbf{P}_n^*(B)$ en este sentido se ofrece en el § 6. Teoremas más exactos sobre la convergencia con probabilidad 1 se dan en el § 4.

§ 3. Características muestrales. Dos tipos de estadísticas

1. Ejemplos de características muestrales. Por *características muestrales* suelen entenderse las diversas funcionales medibles de una distribución empírica o, dicho de otro modo, las funciones de una muestra que se supone que son medibles. Entre ellas, los momentos muestrales (o empíricos) son los más simples. Llámase *momento muestral de orden k* el valor de

$$a_k^* = a_k^*(X) = \int x^k dF_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^k.$$

El momento central muestral de orden k es igual a

$$a_k^{*\circ} = a_k^{*\circ}(X) = \int (x - a_1^*)^k dF_n^*(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - a_1^*)^k.$$

Para los momentos muestrales a_1^* y $a_2^{*\circ}$, en la literatura se utilizan designaciones especiales, $\bar{x}$ y S^2:

$$\bar{x} = a_1^* = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad S^2 = a_2^{*\circ} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

En los problemas estadísticos se usan las características muestrales más diferentes. Por ejemplo, la *mediana muestral* ζ^* es el valor medio de una serie variacional, o sea, el valor de $\zeta^* = x_{(m)}$ si $n = 2m - 1$ (impar) y $\zeta^* = (x_{(m)} + x_{(m+1)})/2$ si $n = 2m$ (par). Recordemos que por mediana ζ de la distribución continua $\mathbf{P}$ se entiende la solución de la ecuación $F(\zeta) = 1/2$.

Un concepto más general es el de *cuantila* ζ_p de orden p. Es el número para el cual $F(\zeta_p) = p$. Así que la mediana es una cuantila de orden 1/2. Si F tiene puntos de discontinuidad (componente discreta) entonces esta definición pierde su sentido. Por eso en un caso general utilizaremos la definición siguiente:

Se denomina *cuantila* ζ_p *de orden* p de la distribución $\mathbf{P}$ el número

$$\zeta_p = \sup \{x: F(x) \leqslant p\}.$$

Como función de p la cuantila ζ_p no es más que la función $F^{-1}(p)$, inversa a $F(x)$.

Es evidente que, a diferencia de la anterior, esta definición de ζ_p (o de $F^{-1}(p)$) tiene sentido para cualesquiera $F(x)$.

Es natural que a la par con las medianas muestrales podemos examinar las *cuantilas muestrales* ζ_p^* *de orden* p que por definición son iguales al valor de $x_{(l)}$, donde $l = [np] + 1$, $x_{(k)}$ son los términos de la serie variacional para la muestra X, $k = 1, \ldots, n$. Para $p = 1/2$ utilizaremos la definición $\zeta^* = \zeta_{1/2}^*$ que hemos dado anteriormente (coincide tan sólo con la definición dada para n impares).

2. Dos tipos de estadísticas. Supongamos que se da una función medible S de n argumentos. La característica muestral $S(X) = S(x_1, \ldots, x_n)$ a menudo también se llama *estadística*. De lo dicho anteriormente se deduce que cualquier estadística es una variable aleatoria. Su distribución se determina por completo mediante la distribución $\mathbf{P}(B) = \mathbf{P}(x_1 \in B)$ (recordemos que $S(X)$ se puede considerar como una variable aleatoria dada en $(\mathscr{X}^n, \mathfrak{B}^n_{\mathscr{X}}, \mathbf{P})$, donde $\mathbf{P}$ es el producto directo múltiplo de n de las distribuciones unidimensionales de x_1).

Destaquemos aquí dos clases de características que se encontrarán frecuentemente a continuación. Se construirán con ayuda de los dos tipos siguientes de funcionales $G(F)$ de las funciones de distribución F:

I. Funcionales que tienen la forma

$$G(F) = h\left(\int g(x)dF(x)\right),$$

donde g es la función dada de Borel; h, la función continua en el punto $a = \int g(x)dF_0(x)$, donde F_0 es tal que $X \in F_0$.

II. Funcionales $G(F)$ continuas en el "punto" F_0 en la métrica uniforme: $G(F^{(n)}) \to G(F_0)$, si $\sup_x |F^{(n)}x) - F_0(x)| \to 0$, los portadores *) de las distribuciones de $F^{(n)}$ pertenecen al portador de F_0. Aquí, como antes, F_0 es la función para la cual $X \in F_0$.

---

*) El portador N_P de la distribución $\mathbf{P}$ con la función de distribución F es el conjunto para el cual $\mathbf{P}(N_P) = 1$.

Vamos a definir las clases respectivas de estadísticas con ayuda de la igualdad

$$S(X) = G(F_n^*),$$

donde F_n^* es la función empírica de distribución. Entonces obtenemos:

I. Clase de *estadísticas de tipo* I, representables en la forma

$$S(X) = h\left(\int g(x)dF_n^*(x)\right) = h\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right).$$

Es evidente que todos los momentos muestrales tienen la forma de las estadísticas aditivas $\frac{1}{n}\sum_{i=1}^{n} g(x_i)$ y figuran entre las estadísticas del tipo I.

II. Clase de estadísticas que llamaremos *estadísticas de tipo* II o bien estadísticas *continuas en el punto* F_0.

Está claro que, por ejemplo, la mediana muestral será la estadística continua en el punto F si existe la mediana ζ, $F(\zeta) = 1/2$ y F es continua y crece estrictamente en el punto ζ.

La pertenencia de las funcionales a una de las clases mencionadas no es, desde luego, alternativa. La funcional $G(F)$ puede no pertenecer a ninguna de estas clases o pertenecer a ambas clases a la vez. Por ejemplo, si G es una funcional de tipo I, el portador de F está concentrado en el segmento $[a, b]$ ($F(a) = 0$, $F(b) = 1$) y la función g tiene una variación limitada en $[a, b]$, entonces G será simultáneamente una funcional de tipo II, ya que en este caso la funcional

$$\int g(x)dF(x) = g(b) - \int_a^b F(x)dg(x)$$

es continua con respecto a F en la métrica uniforme. Lo dicho quiere decir que las estadísticas de tipo I $\bar{x}$ y S^2 serán también de tipo II si $X \in \mathbf{P}$ y $\mathbf{P}$ está concentrada en el intervalo finito.

Podemos completar los teoremas 2.1 y 2.2 con la siguiente afirmación sobre la convergencia casi segura de las características muestrales.

Teorema 1. *Sea, como antes,* $X_n = [X_\infty]_n \in F$. *En este caso, si* $S(X) = G(F_n^*)$ *es la estadística de tipo* I ó II, *para* $n \to \infty$

$$G(F_n^*) \xrightarrow[c.s.]{} G(F).$$

Aquí se supone, desde luego, que el valor de $G(F)$ existe.

Ahora bien, las muestras de gran volumen permiten estimar no sólo la propia distribución $\mathbf{P}$, sino también las funcionales de esta distribución,

por lo menos aquellas que pertenecen a una de las clases citadas en el teorema.

Demostración de la afirmación para ambas clases de estadísticas es casi evidente. Sea, por ejemplo, $G(F) = h\left(\int g(x)dF(x)\right)$. Entonces

$$S = S(X) = \int g(x)dF_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$$

es la suma de las variables aleatorias independientes, con la esperanza matemática

$$\mathbf{M}g(x_1) = \int g(x)dF(x).$$

Por eso en consonancia con la ley fuerte de los grandes números $S \underset{c.s.}{\longrightarrow} \mathbf{M}g(x_1)$. Sea ahora $A = \{X_\infty: S(X) \to \mathbf{M}g(x_1)\}$. Entonces $\mathbf{P}(A) = 1$ y si $X_\infty \in A$, entonces $S(X) \to \mathbf{M}g(x_1)$, $h(S(X)) \to h(\mathbf{M}g(x_1))$. Con otras palabras, en el conjunto A

$$G(F_n^*) \to G(F).$$

La afirmación del teorema para las funcionales de segundo tipo es el corolario directo del teorema de Glivenko — Cantelli. ◁

Del teorema se deduce que los momentos absolutos y centrales convergen casi seguramente para $n \to \infty$ a los momentos correspondientes de la distribución $\mathbf{P}$:

$$a_k^* = a_k^*(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^k \underset{c.s.}{\longrightarrow} \mathbf{M}x_1^k,$$

$$a_k^{*\circ} = a_k^{*\circ}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k \underset{c.s.}{\longrightarrow} \mathbf{M}(x_1 - \mathbf{M}x_1)^k.$$

En particular,

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2 \to \mathbf{D}x_1.$$

Ahora bien, hemos establecido un hecho importante que tiene para nosotros el valor de principio: con el aumento del volumen de la muestra, la distribución empírica y una amplia clase de funcionales de ésta se aproximan indefinidamente a los valores "teóricos" correspondientes.

Teoremas más exactos de la distribución de las características muestrales se exponen en los §§ 7 y 8.

§ 4. Muestras multidimensionales

1. Distribuciones empíricas. De un modo completamente análogo se construyen las distribuciones empíricas y las características muestrales en el caso multidimensional cuando la variable aleatoria observada ξ, y junto con ella también los valores muestrales $x_1, \ldots, x_n$, son vectores de dimensión $m > 1$: $x_k = (x_{k,1}, \ldots, x_{k,m})$. Aquí $\mathbf{P}(B) = \mathbf{P}(\xi \in B)$ es la distribución en $\mathscr{X} = R^m$, y el espacio muestral aquí será $(\mathscr{X}^n, \mathfrak{B}_{\mathscr{X}}^n, \mathbf{P})$, donde $\mathbf{P}$ es el producto directo múltiplo de n de las distribuciones $\mathbf{P}$ en $(R^m, \mathfrak{B}_{\mathscr{X}} = \mathfrak{B}_R^m)$. La designación $X \Subset \mathbf{P}$ conserva por completo su sentido.

La distribución empírica $\mathbf{P}_n^*$, basada en la muestra X, se construye, al igual que antes, como una distribución discreta con masas de valores $1/n$ en los puntos $x_1, \ldots, x_n$, así que

$$\mathbf{P}_n^*(B) = \frac{\nu(B)}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_{x_i}(B),$$

donde $\nu(B)$ es el número de puntos que entran en el conjunto B, y $\mathbf{I}_{x_i}$, la distribución concentrada en el punto x_i.

Es evidente que la afirmación del teorema 1 acerca de la convergencia de $\mathbf{P}_n^*(B) \xrightarrow[c.s.]{} \mathbf{P}(B)$ aquí también será válida.

La generalización del teorema de Glivenko — Cantelli para el caso multidimensional está relacionada con la aparición de cuestiones cualitativamente nuevas. Una de ellas consiste en generalizar el concepto de intervalos para el caso multidimensional. Puede haber varias generalizaciones de tal género, por ejemplo, rectángulos, conjuntos convexos, etc.

Una variante elemental de generalización del teorema de Glivenko — Cantelli es la siguiente.

Sea $y = (y_1, \ldots, y_m)$ el punto R^m, y B_t, un ángulo con vértice en el punto $t = (t_1, \ldots, t_m)$:

$$B_t = \{y \in R^m: \quad y_k < t_k, \quad k = 1, \ldots, m\}.$$

La función

$$F_n^*(t) = \mathbf{P}_n^*(B_t)$$

se llama *función empírica de distribución.*

Teorema 1. *Sea* $X_n = [X_\infty]_n$, $X_\infty \Subset F$. *Entonces*

$$\sup_t |F_n^*(t) - F(t)| \xrightarrow[c.s.]{} 0$$

si $n \to \infty$.

2*. Variantes más generales del teorema de Glivenko — Cantelli. Ley de logaritmo repetido. Una de las generalizaciones posibles de los teoremas

del tipo de Glivenko — Cantelli consiste en lo siguiente. Sea $\mathfrak{C}$ la clase de todos los conjuntos convexos sobre R^m.

Teorema 2. *Supongamos que $X_n = [X_\infty]_n$, $X_\infty \in \mathbf{P}$ y que la distribución $\mathbf{P}$ es absolutamente continua respecto a la medida de Lebesgue en R^m. Entonces*

$$\sup_{B \in \mathfrak{C}} |\mathbf{P}_n^*(B) - \mathbf{P}(B)| \underset{c.s.}{\to} 0. \tag{1}$$

Otras generalizaciones posibles del teorema 1 pueden ser obtenidas con ayuda de las afirmaciones del Suplemento 1.

Observación 1. La exigencia de que la distribución $\mathbf{P}$ sea absolutamente continua con respecto a la medida de Lebesgue es muy importante en el teorema 2. Esto lo demuestra el ejemplo siguiente. Sea $\mathbf{P}$ la distribución uniforme en una circunferencia unitaria (o sea, en el límite de un círculo) en R^2. Construyamos el polígono cerrado B_X con los vértices en los puntos $x_1, \dots, x_n$ situados en dicha circunferencia. Es un conjunto convexo. Sin embargo, $\mathbf{P}(B_X) = 0$, $\mathbf{P}_n^*(B_X) = 1$, es incorrecta y, por consiguiente, también lo es la relación (1), donde $\mathfrak{C}$ es la clase de los conjuntos convexos.

Las afirmaciones de los teoremas del tipo de Glivenko — Cantelli pueden ser precisadas considerablemente, por lo menos, para las clases elementales de conjuntos. Por ejemplo, para las funciones empíricas de distribuciones $F_n^*(t)$ (véase el teorema 1) se puede señalar la siguiente sucesión determinada: $b_n \to 0$ cuando $n \to \infty$, para la cual, con la probabilidad 1 (para casi todos los "puntos" X_∞),

$$\varlimsup_{n\to\infty} b_n^{-1} \sup_t |F_n^*(t) - F(t)| = 1.$$

Resulta que el orden de pequeñez de b_n equivale al de $\sqrt{\dfrac{\ln \ln n}{n}}$.

Teorema 3 (ley del logaritmo repetido). *Si $F(t)$ es continua, entonces*

$$\mathbf{P}\left(\varlimsup_{n\to\infty} \sqrt{\frac{2n}{\ln \ln n}} \sup_t |F_n^*(t) - F(t)| = 1\right) = 1.$$

El teorema 3 está estrechamente relacionado con la aproximación normal para $F_n^*(t)$ de la forma (2.8) que, evidentemente, en el caso multidimensional también tiene lugar.

La demostración de los teoremas 1 y 2 se da en el Suplemento 1, y la demostración del teorema 3 véase en [52].

3. Características muestrales. En el caso multidimensional, al igual que en el unidimensional, éstas son distintas funciones medibles de la muestra. Las más elementales de ellas son los momentos muestrales. Por ejemplo, los momentos muestrales de primer orden son iguales a

$$a_{1j}^* = a_{1j}^*(X) = \frac{1}{n}\sum_{k=1}^{n} x_{k,j}, \quad j = 1, \dots, m.$$

Los momentos de segundo orden (ordinarios y centrales)

$$a_{2,ij}^* = a_{2,ij}^*(X) = \frac{1}{n}\sum_{k=1}^{n} x_{k,i}x_{k,j}; \quad i, j = 1, \dots, m,$$

$$a_{2,ij}^{*0} \equiv S_{ij} = \frac{1}{n}\sum_{k=1}^{n} (x_{k,i} - a_{1i}^*)(x_{k,j} - a_{1j}^*),$$

etc. Al igual que en el caso unidimensional, con ayuda de la ley fuerte de los grandes números es fácil cerciorarse de que estas características convergen, con probabilidad 1, hacia los momentos "teóricos" correspondientes. En particular, $S_{ij} \underset{c.s.}{\to} \mathbf{M}(x_{1,i} - \mathbf{M}x_{1,i})(x_{1j} - \mathbf{M}x_{1j})$. Es fácil convencerse (esto se analiza más detalladamente en el párrafo siguiente) de que los coeficientes de correlación muestrales

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} \underset{c.s.}{\longrightarrow} \varrho(x_{1,i}x_{1,j}) = \frac{\mathbf{M}(x_{1,i} - \mathbf{M}x_{1,i})(x_{1j} - \mathbf{M}x_{1j})}{\sqrt{\mathbf{D}x_{1,i}\mathbf{D}x_{1j}}}$$

también poseen esta misma propiedad.

Para obtener teoremas más exactos de la distribución de las características muestrales nos serán útiles los llamados teoremas de continuidad.

§ 5. Teoremas de continuidad

En lo sucesivo necesitaremos ciertos conceptos auxiliares que utilizaremos a menudo y que podrían ser llamados *teoremas de continuidad*. Para facilitar su estudio, a ellos les dedicamos un párrafo especial. Anteriormente ya hemos utilizado un teorema de este tipo — el teorema 3.1. El primer teorema de continuidad será muy parecido a éste.

Teorema 1 (primer teorema de continuidad). *Sea $X = |X_\infty|_n \in \mathbf{P}$. En este caso, si $S_n = S_n(X)$ es una sucesión de estadísticas escalares o vectoriales, tales que $S_n \underset{c.s.}{\to} S_0$, y $H(s)$ es una función continua casi por doquier con respecto a la distribución de la variable aleatoria S_0 (o sea, $H(s)$ es continua en cada punto del conjunto B, $\mathbf{P}(S_0 \in B) = 1$), entonces $H(S_n(X)) \underset{c.s.}{\longrightarrow} H(S_0)$.*

Si S_n converge hacia S_0 según la probabilidad ($S_n \underset{p}{\to} S_0$), entonces para las demás condiciones semejantes, $H(S_n) \underset{p}{\to} H(S_0)$.

La demostración del teorema es casi evidente. Como las probabilidades de los sucesos $A = \{X_\infty : S_n(X_\infty) \to S_0(X_\infty)\}$ y $C = \{X_\infty : S_0(X_\infty) \in B\}$ son iguales a 1, entonces, en virtud de la igualdad $\mathbf{P}(A \cap C) = \mathbf{P}(A) + \mathbf{P}(C) - \mathbf{P}(A \cup C)$, la probabilidad del suceso $A \cap C$ (en el cual $H(S_n(X_\infty)) \to H(S_0(X_\infty))$) también es igual a 1.

Para simplificar la demostración de la convergencia en probabilidad, supongamos adicionalmente que $S_0 =$ const (sólo necesitaremos este caso). Para un valor dado de $\varepsilon > 0$ hay un valor de $\delta > 0$ tal, que el suceso $A_n = \{X_\infty : |S_n - S_0| < \delta\}$ contribuye a que $|H(S_n) - H(S_0)| < \varepsilon$ y además, $\mathbf{P}(A_n) > 1 - \varepsilon$ para todos los valores de n bastante grandes. Por lo tanto, para tales n tenemos $1 - \varepsilon < \mathbf{P}(A_n) \leqslant \mathbf{P}(|H(S_n) - H(S_0)| < \varepsilon)$. ◅

Antes de enunciar los teoremas siguientes, introduzcamas ciertas designaciones que serán cómodas posteriormente.

Supongamos que se ha dado una sucesión de vectores aleatorios $\eta_n = (\eta_n^{(1)}, \ldots, \eta_n^{(s)})$ (no obligatoriamente en el mismo espacio probabilístico). Si las distribuciones η_n convergen débilmente (cuando $n \to \infty$) hacia la distribución de cierta variable aleatoria η, entonces designaremos este hecho con el símbolo

$$\eta_n \Rightarrow \eta. \tag{1}$$

Aquí utilizamos, para las variables aleatorias, el signo $\Rightarrow$ de convergencia débil de las distribuciones. Al igual que antes, utilizaremos también este signo para las propias distribuciones, así que la relación (1) es equivalente a que

$$\mathbf{Q}_n \Rightarrow \mathbf{Q},$$

donde $\mathbf{Q}_n$ y $\mathbf{Q}$ son las distribuciones de η_n y η respectivamente. Tal convenio es cómodo y no conduce a equivocaciones.

Está claro que de $\eta_n \xrightarrow[p]{} \eta$ o de $\eta_n \xrightarrow[c.s.]{} \eta$ se deduce $\eta_n \Rightarrow \eta$ (compárese con [11], p. 133).

Ahora bien, si se trata de la relación (correspondiente a una convergencia débil) entre objetos de igual naturaleza (entre variables aleatorias o entre distribuciones), usaremos el símbolo $\Rightarrow$. También sería conveniente tener el símbolo para expresar el hecho de que "*las distribuciones de* η_n *convergen débilmente hacia* $\mathbf{Q}$ *cuando* $n \to \infty$". Escribiremos esta relación de la forma

$$\eta_n \Subset \mathbf{Q}, \tag{2}$$

así que el símbolo $\Subset$ expresa el mismo hecho que $\Rightarrow$, pero une objetos de distinta naturaleza, al igual que el símbolo $\in$ respecto a $\eta \Subset \mathbf{Q}$ (a la izquierda en (2) se encuentran las variables aleatorias, y a la derecha, la distribución).

Sean η_n y η vectores aleatorios de R^s.

Teorema 2 (segundo teorema de continuidad). *Si* $\eta_n \Rightarrow \eta$ *y* $H(t)$, $t \in R^s$ *es una función continua de* R^s *en* R^k, *entonces* $H(\eta_n) \Rightarrow H(\eta)$.

Señalemos que, en realidad, este teorema también es cierto en una forma más general [*)]. *Si* $\eta_n \Rightarrow \eta$ *y* $H(t)$ *es continua en los puntos del conjunto* $A \in \mathfrak{B}^s$, $\mathbf{P}(\eta \in A) = 1$, *entonces* $H(\eta_n) \Rightarrow H(\eta)$.

Demostración del teorema 2. Sean $\mathbf{Q}_n$ y $\mathbf{Q}$ las distribuciones η_n y η, respectivamente. La convergencia débil de $\mathbf{Q}_n \Rightarrow \mathbf{Q}$ significa, por definición, que para toda función continua y limitada f: $R^s \to R$ se cumple

$$\int f(y)\mathbf{Q}_n(dy) \to \int f(y)\mathbf{Q}(dy)$$

o bien, que es lo mismo,

$$\mathbf{M}f(\eta_n) \to \mathbf{M}f(\eta). \tag{3}$$

También debemos obtener una relación análoga para las distribuciones $H(\eta_n)$ y $H(\eta)$. O sea, debemos establecer que para toda función continua limitada $g: R^k \to R$ es válida $\mathbf{M}g(H(\eta_n)) \to \mathbf{M}g(H(\eta))$. Pero esto se deduce con evidencia de (3), ya que la superposición $\tilde{g} = g \circ H$: $R^s \to R$ es continua y limitada. ◁

Teorema 3 (tercer teorema de continuidad). *Sea* $\eta_n \Rightarrow \eta \in R$, $H(t)$, $t \in R$ *una función derivable en el punto* a. *Entonces, si* $b_n \to 0$ *es una sucesión numérica,*

$$(H(a + b_n\eta_n) - H(a))/b_n \Rightarrow \eta H'(a). \tag{4}$$

Demostración. Examinemos la función

$$h(x) = \begin{cases} (H(a + x) - H(a))/x, & x \neq 0, \\ H'(a), & x = 0, \end{cases}$$

la cual será continua en el punto $x = 0$. Como $b_n\eta_n \Rightarrow 0$, en virtud del primer teorema de continuidad, $h(b_n\eta_n) \Rightarrow h(0) = H'(a)$. Utilizando el segundo teorema de continuidad, obtenemos

$$(H(a + b_n\eta_n) - H(a))/b_n = h(b_n\eta_n)\eta_n \Rightarrow H'(a)\eta. \quad ◁$$

Ahora citaremos dos generalizaciones sucesivas del teorema 3 para el caso multidimensional, las cuales nos serán útiles.

Teorema 3A. *Supongamos que* $\eta_n = (\eta_n^{(1)}, \ldots, \eta_n^{(s)}) \Rightarrow \eta = (\eta^{(1)}, \ldots, \eta^{(s)})$ *y que* $H(t)$ *es función escalar del vector* $t = (t_1, \ldots, t_s)$ *con la que existe la derivada* $H'(t) = \left(\frac{\partial H}{\partial t_1}, \ldots, \frac{\partial H}{\partial t_s}\right)$ *en el punto* a. *Entonces, cuando* $b_n \to 0$,

---

[*)] Véase [5].

$$(H(a + b_n\eta_n) - H(a))/b_n \Rightarrow \eta(H'(a))^T = \sum_{j=1}^{s} \frac{\partial H(a)}{\partial t_j} \eta^{(j)}. \quad (5)$$

Aquí el índice T corresponde a la transposición.

Si $\eta(H'(a))^T = 0$ *con probabilidad* 1 *(por ejemplo,* $H'(a) = 0$*), y la matriz* $H''(t)$ *de las derivadas* $\frac{\partial^2 H(t)}{\partial t_i \partial t_j}$ *existe en el punto a, entonces*

$$(H(a + b_n\eta_n) - H(a))/b_n^2 \Rightarrow \frac{1}{2}\eta H''(a)\eta^T = \frac{1}{2}\sum_{i,j=1}^{s} \frac{\partial^2 H(a)}{\partial t_i \partial t_j} \eta^{(i)}\eta^{(j)}. \quad (6)$$

Sea ahora $H(t)$ una función vectorial. Entonces, evidentemente, la distribución límite para cada componente H_j será descrita por el teorema 3A, y con respecto a la distribución conjunta será valida.

Teorema 3B. *Supongamos que* $\eta_n \Rightarrow \eta \in R^s$ *y que* $H(t) \in R^k$ *es una función vectorial con la que las derivadas* H_j', $j = 1, \dots, k$ *satisfacen las condiciones del teorema* 3A. *Entonces*

$$(H(a + b_n\eta_n) - H(a))/b_n \Rightarrow \eta(H'(a))^T.$$

Si $\eta(H'(a))^T = 0$ *con probabilidad* 1, *y las matrices* H_j'', $j = 1, \dots, k$ *existen en el punto a, entonces*

$$(H(a + b_n\eta_n) - H(a))/b_n^2 \Rightarrow \frac{1}{2}(\eta H_1''(a)\eta^T, \dots, \eta H_k''(a)\eta^T).$$

Las demostraciones de estas afirmaciones, de hecho no se distinguen en nada de la demostración del teorema 3, y por eso las presentamos al lector en calidad de ejercicios. Además, proponemos convencerse de que el símbolo $\Rightarrow$ en (4)—(6) se puede sustituir por $\underset{c.s}{\longrightarrow}$ o por $\underset{P}{\rightarrow}$, si se cumple $\eta_n \underset{c.s}{\rightarrow} \eta$ o $\eta_n \underset{P}{\rightarrow} \eta$, respectivamente.

El contenido de los teoremas 1—3 puede resumirse del modo siguiente. Supongamos que $\sim\rightarrow$ significa uno de los símbolos $\underset{c.s}{\longrightarrow}$, $\underset{P}{\rightarrow}$, $\Rightarrow$. Entonces, si H es continua, de $\eta_n \sim\rightarrow \eta$ resulta $H(\eta_n) \sim\rightarrow H(\eta)$.

Si H es derivable en el punto a, $\eta_n \sim\rightarrow \eta$, entonces para $b_n \rightarrow 0$

$$(H(a + b_n\eta_n) - H(a))/b_n \sim\rightarrow H'(a)\eta. \quad (7)$$

Observación 1. No es difícil notar que si a depende de n de modo que $a \equiv a_n = a_0 + o(1)$ y las derivadas en los teoremas 3, 3A y 3B son continuas, la relación (7) se conservará en la forma

$$(H(a_n + b_n\eta_n) - H(a_n))/b_n \sim\rightarrow H'(a_0)\eta. \quad (8)$$

Para la demostración es suficiente ver que el primer miembro (8) es representable en forma de $H'(\alpha_n)\eta_n$, donde $\alpha_n = \theta a_n + (1-\theta)(a_n + b_n\eta_n) \sim \to a_0$, $|\theta| \leqslant 1$, y utilizar el segundo teorema de continuidad.

Esa misma observación es válida para los análogos multidimensionales de la referida afirmación en los teoremas 3A, 3B.

Los teoremas enunciados conciernen a la convergencia casi segura y a la convergencia de las distribuciones. El cuarto teorema de continuidad se refiere a la convergencia de las integrales.

Teorema 4 (teorema de continuidad para los momentos). *Supongamos que* $\{\eta_n\}$ *es una sucesión de variables aleatorias numéricas y que* $\eta_n \Rightarrow \eta$ *cuando* $n \to \infty$. *En este caso, si se cumple al menos una de las condiciones siguientes*:

1) $\varlimsup_{n\to\infty} \int_N^\infty \mathbf{P}(|\eta_n| > x)dx \to 0$ *para* $N \to \infty$,

2) $\mathbf{P}(|\eta_n| > x) \leqslant \varphi(x)$, $\int_0^\infty \varphi(x)dx < \infty$,

3) $\mathbf{M}|\eta_n|^{1+\alpha} < c < \infty$ para cierto $\alpha > 0$,

entonces $\lim_{n\to\infty} \mathbf{M}\eta_n = \mathbf{M}\eta$.

Nótese que la condición 1 significa la convergencia uniforme en n hacia el cero $\int_N^\infty \mathbf{P}(|\eta_n| > x)dx$ cuando $N \to \infty$.

Demostración. De la desigualdad generalizada de Chébishev,

$$\mathbf{P}(|\eta_n| > x) \leqslant \frac{\mathbf{M}|\eta_n|^{1+\alpha}}{x^{1+\alpha}},$$

se deduce que la condición 3 provoca la condición 2 y ésta, a su vez, la condición 1.

Supongamos que se ha cumplido la condición 1. Para simplificar los razonamientos, admitamos primeramente que $\eta_n \geqslant 0$. Entonces, integrando por partes, obtenemos

$$\mathbf{M}\eta_n = -\int_0^\infty x\, d\mathbf{P}(\eta_n \geqslant x) = \int_0^\infty \mathbf{P}(\eta_n \geqslant x)dx.$$

De esta representación, así como de la convergencia de $\mathbf{P}(\eta_n \geqslant x) \to \mathbf{P}(\eta \geqslant x)$ para casi todos los x, y de la convergencia, uniforme en n, de la integral $\int_0^\infty \mathbf{P}(\eta_n \geqslant x)dx$, se deduce la legitimidad del paso límite bajo el signo de integral, en virtud del cual

$$\lim_{n\to\infty} \mathbf{M}\eta_n = \lim_{n\to\infty} \int_0^\infty \mathbf{P}(\eta_n \geqslant x)dx = \int_0^\infty \mathbf{P}(\eta \geqslant x)dx = \mathbf{M}\eta.$$

En el caso general conviene utilizar la representación $\eta_n = \eta_n^+ - \eta_n^-$, donde $\eta_n^+ = \text{máx}(\eta_n, 0)$, $\eta_n^- = \text{máx}(-\eta_n, 0)$. ◁

Señalemos que la condición 1 también puede considerarse como condición de la integrabilidad uniforme de η_n, de la cual se deduce inmediatamente la convergencia requerida de $\mathbf{M}\eta_n \to \mathbf{M}\eta$ (véase, por ejemplo, [11], [60]).

§ 6*. Función empírica de distribución como proceso aleatorio. Convergencia hacia el puente browniano

En este párrafo supondremos que se conoce el concepto de *proceso aleatorio* (digamos, en el volumen de [11]) y, en particular, las definiciones y propiedades elementales de los *procesos wieneriano y poissoniano*.

1. Distribución del proceso $nF_n^*(t)$. Nos limitaremos a examinar el caso unidimensional $\mathscr{X} = R$. Supongamos, como antes, que $F_n^*(t) = \mathbf{P}_n^*((-\infty, t))$ es la función empírica de distribución correspondiente a la muestra $X = X_n \in \mathbf{P}$.

La función $F_n^*(t)$ es una función de dos variables: t y X, o bien que es lo mismo, una función aleatoria de t o un *proceso aleatorio*.

Hallemos las *distribuciones de dimensión finita* de este proceso. Supongamos $t_1 < t_2 < \ldots < t_m$ son m puntos arbitrarios del eje numérico. Pongamos $t_0 = -\infty$, $t_{m+1} = \infty$ y designemos por

$$\Delta_j g = g(t_{j+1}) - g(t_j)$$

los incrementos de la función $g(t)$ en los semiintervalos $\Delta_j = [t_j, t_{j+1})$, $j = 0, 1, \ldots, m$. Examinemos el incremento $\Delta_j\pi_n$ del proceso

$$\pi_n(t) = nF_n^*(t).$$

Evidentemente, esto es el número de elementos de la muestra que se encuentran en Δ_j. La probabilidad de que un elemento de la muestra (digamos, x_1) se halle en Δ_j es igual a $p_j = \mathbf{P}(\Delta_j)$. Como el hecho de que los elementos tomen un valor perteneciente a Δ_j, $j = 0, 1, \ldots, m$, constituye $m + 1$ sucesos incompatibles, tenemos aquí, sin duda, una distribución polinomial (véase [11], p. 111) para el vector $(\Delta_0\pi_n, \ldots, \Delta_m\pi_n)$ con probabilidades $p_0, \ldots, p_m$, $\sum_{j=0}^{m} p_j = 1$. Como es sabido,

$$\mathbf{P}(\Delta_0\pi_n = k_0, \ldots, \Delta_m\pi_n = k_m) = \frac{n!}{k_0! \ldots k_m!} p_0^{k_0} \ldots p_m^{k_m}, \qquad (1)$$

donde $\sum_{j=0}^{m} k_j = n$.

Sea ahora $\eta(u)$, $u \in [0, 1]$, el *proceso poissoniano* continuo a la izquierda (véase [11], p. 304) *con parámetro* λ, $\eta(0) = 0$. Los incrementos de este proceso son independientes,

$$\mathbf{P}(\eta(u) = k) = e^{-\lambda u} \frac{(\lambda u)^k}{k!}.$$

Si la función de distribución $F(t) = \mathbf{P}((-\infty, t))$ es continua, podemos realizar la sustitución continua del tiempo, poniendo $u = F(t)$, $-\infty < t < \infty$, y determinar de este modo el proceso $\pi(t) = \eta(F(t))$ sobre todo el eje. Examinemos los incrementos de este proceso

$$\Delta_j\pi = \pi(t_{j+1}) - \pi(t_j) = \eta(F(t_{j+1})) - \eta(F(t_j))$$

sobre los intervalos Δ_j. Entonces

$$\mathbf{P}(\Delta_0\pi = k_0, \ldots, \Delta_m\pi = k_m) = \prod_{j=0}^{m} e^{-\lambda p_j} \frac{(\lambda p_j)^{k_j}}{k_j!} = e^{-\lambda}\lambda^n \prod_{j=0}^{m} \frac{p_j^{k_j}}{k_j!}$$

y la probabilidad condicional de este mismo proceso, a condición de que $\pi(\infty) = \sum_{j=0}^{m} \Delta_j\pi = n$, será igual a

$$\mathbf{P}\left(\Delta_0\pi = k_0, \ldots, \Delta_m\pi = k_m \,\middle|\, \sum_{j=0}^{m} \Delta_j\pi = n\right) =$$

$$= \frac{\mathbf{P}(\Delta_0\pi = k_0, \ldots, \Delta_m\pi = k_m)}{\mathbf{P}(\pi(\infty) = n)} =$$

$$= \mathbf{P}(\Delta_0\pi = k_0, \ldots, \Delta_m\pi = k_m) \frac{e^{\lambda} n!}{\lambda^n} = n! \prod_{j=0}^{m} \frac{p_j^{k_j}}{k_j!}. \quad (2)$$

Hemos obtenido para cualquier $\lambda > 0$ la misma expresión que en el segundo miembro de (1). Así pues, hemos demostrado la afirmación siguiente.

Teorema 1. *Si $F(t)$ es continua, la distribución del proceso $nF_n^*(t)$ coincide con la distribución condicional del proceso $\pi(t) = \eta(F(t))$ a condición de que $\pi(\infty) = n(\eta(1) = n)$.*

El teorema muestra que las desviaciones $n(F_n^*(t) - F(t))$ están distribuidas al igual que $\eta(F(t)) - nF(t)$ a condición de que $\eta(1) = n$ y el problema con precisión hasta la sustitución del tiempo $u = F(t)$ se reduce al estudio de las desviaciones $\eta(u) - nu$ para el proceso poissoniano condicional ($\eta(1) = n$) sobre el segmento [0, 1] o bien, que es lo mismo, al estudio

de las desviaciones $n(F_n^*(t) - t)$, donde $F_n^*(t)$ corresponde a la distribución uniforme sobre $[0, 1]$.

Puede ser útil también otra representación para el proceso $nF_n^*(t)$. Sean $\zeta_1, \zeta_2, \ldots$ los puntos de saltos del proceso poissoniano $\eta(t)$, así que $\eta(\zeta_k + 0) = k$. Como es sabido ([11]), las diferencias $\xi_k = \zeta_k - \zeta_{k-1}$ ($\zeta_0 = 0$), $k = 1, 2, \ldots$, son independientes y están distribuidas exponencialmente

$$\mathbf{P}(\xi_k > x) = e^{-\lambda x},$$

ζ_k tiene Γ-distribución con densidad (véase también el § 2.2)

$$\gamma_{\lambda,k}(x) = \frac{\lambda^k}{\Gamma_{(k)}} e^{-\lambda x} x^{k-1}.$$

Para simplificar las enunciaciones, supongamos que $F(t) = t$, $t \in [0, 1]$, $t_0 = 0$, $t_{m+1} = 1$, así que $\eta(t) = \pi(t)$.

Teorema 2. *La distribución del proceso $nF_n^*(t)$ coincide, para cualquier $v > 0$, con la distribución condicional del proceso $\pi(tv)$, $0 < t < 1$, a condición de que $\zeta_{n+1} = v$.*

Con otras palabras, la afirmación del teorema 1 seguirá válida si la condición $\pi(1) = n$ se sustituye por una condición mucho más estrecha $\pi(1) = n$, $\pi(1 + 0) = n + 1$ (suponemos que las trayectorias de $\pi(t)$ son continuas a la izquierda).

Como la probabilidad de esta nueva condición es igual a 0, puede ser que convenga añadir (véanse los §§ 4 y 8 en [11] sobre las esperanzas matemáticas, así como el § 2.9) que por distribución condicional entendemos las probabilidades

$$\mathbf{P}(A/\zeta_{n+1} = v) = \frac{\mathbf{P}(A;\ \zeta_{n+1} \in dv)}{\mathbf{P}(\zeta_{n+1} \in dv)},$$

donde $A = \{\Delta_0\pi(tv) = k_0, \ \ldots, \ \Delta_m\pi(tv) = k_m\}$, $\Delta_j\pi(tv) = \pi(t_{j+1}v) - \pi(t, v)$.

Demostración. Representemos el suceso $\{\zeta_{n+1} \in dv\}$ en la forma del producto de dos sucesos

$$B = \{\pi(v) = n\} \text{ y } C = \{\pi(v + dv) - \pi(v) = 1\}.$$

Los sucesos B y AB no dependen de C, ya que los sucesos B y AB, por un lado, y el suceso C, por otro, se refieren a los incrementos del proceso π sobre los intervalos disjuntos del tiempo. Por eso

$$\mathbf{P}(A/\zeta_{n+1} = v) = \frac{\mathbf{P}(ABC)}{\mathbf{P}(BC)} = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)} = \mathbf{P}(A/\pi(v) = n). \quad (3)$$

Lo mismo que en (2) nos cercioramos de que esta expresión no depende de v (ni tampoco de λ) y coincide con (1). $\triangleleft$

Corolario 1. *La distribución del proceso $nF_n^*(t)$ coincide con la distribución $\pi(t\zeta_{n+1})$, $0 \leqslant t \leqslant 1$.*

Esto se deduce del hecho de que para $B = \{\Delta_0\pi(t\zeta_{n+1}) = k_0, \ldots$ $\ldots, \Delta_m\pi(t\zeta_{n+1}) = k_m\}$ tenemos, en virtud de (3),

$$P(B) = \int \mathbf{P}(A/\zeta_{n+1} = v)\, \mathbf{P}(\zeta_{n+1} \in dv) = n! \prod_j \frac{\Delta_j^{k_j}}{k_j!}.$$

Del corolario 1 se deduce:

Corolario 2. *La distribución conjunta de los elementos de la serie variacional $x_{(1)}, \ldots, x_{(n)}$ de la muestra X de la distribución uniforme coincide con la distribución conjunta*

$$\frac{\zeta_1}{\zeta_{n+1}}, \ldots, \frac{\zeta_n}{\zeta_{n+1}},$$

o bien, que es lo mismo, la distribución conjunta de las diferencias $x_{(1)}$, $x_{(2)} - x_{(1)}, \ldots, x_{(n)} - x_{(n-1)}$, $1 - x_{(n)}$ coincide con la distribución conjunta

$$\frac{\xi_1}{\zeta_{n+1}}, \ldots, \frac{\xi_{n+1}}{\zeta_{n+1}}.$$

Para concluir este apartado determinaremos los momentos de segundo orden para los incrementos del proceso $n(F_n^*(t) - F(t))$. Para nosotros será más cómodo examinar el proceso

$$w^n(t) = \sqrt{n}(F_n^*(t) - F(t)).$$

Es evidente que $\mathbf{M}\Delta_j w^n = 0$, $\mathbf{M}(\Delta_j w^n)^2 = \Delta_j F(1 - \Delta_j F)$. Para calcular los momentos mixtos notemos que ($i \neq j$)

$$\mathbf{M}(\Delta_i w^n \cdot \Delta_j w^n) = \frac{1}{n} \sum_{k,l=1}^{n} \mathbf{M}(\mathbf{I}_{x_k}(\Delta_i) - \mathbf{P}(\Delta_i)) \times$$

$$\times (\mathbf{I}_{x_l}(\Delta_j) - \mathbf{P}(\Delta_j)) = \frac{1}{n} \sum_{k,\ l=1}^{n} \{\mathbf{M}\mathbf{I}_{x_k}(\Delta_i)\mathbf{I}_{x_l}(\Delta_j) - \mathbf{P}(\Delta_i)\mathbf{P}(\Delta_j)\}.$$

Puesto que

$$\mathbf{M}\mathbf{I}_{x_k}(\Delta_i)\mathbf{I}_{x_l}(\Delta_j) = \begin{cases} \mathbf{P}(\Delta_i)\mathbf{P}(\Delta_j) \text{ para } k \neq l \\ 0 \text{ para } k = l, \end{cases}$$

Entonces $\mathbf{M}(\Delta_i w^n \cdot \Delta_j w^n) = -\mathbf{P}(\Delta_i)\mathbf{P}(\Delta_j) = -\Delta_i F \cdot \Delta_j F.$

Ahora bien, los incrementos del proceso w^n están correlacionados negativamente.

2. Comportamiento límite del proceso $w^n(t)$. Supongamos que $F(t)$ es continua. Del punto 1 entonces se deduce que podemos limitarnos a examinar la distribución $F(t) = t$ uniforme sobre $[0, 1]$, $0 \leqslant t \leqslant 1$.

Designemos por $w(t)$ el *proceso wieneriano estándar*, o sea, el proceso con incrementos independientes para el cual $w(t)$ está distribuido normalmente con parámetros $(0, t)$. El proceso

$$w^\circ(t) = w(t) - tw(1)$$

se llama *puente browniano* (puesto que en él se hallan asegurados ambos extremos: $w^\circ(0) = w^\circ(1) = 0$). La distribución de este proceso coincide con la distribución condicional del proceso $w(t)$ a condición de que $w(1) = 0$ (mejor dicho, es necesario adoptar la condición $|w(1)| < \varepsilon$ y pasar al límite para $\varepsilon \to 0$).

Resulta que las distribuciones de dimensión finita de los procesos

$$w^n(t) = \sqrt{n}(F_n^*(t) - F(t)), \quad t \in [0, 1],$$

convergen, cuando $n \to \infty$, hacia las distribuciones correspondientes del puente browniano $w^\circ(t)$.

Este hecho permite aproximar los procesos $w^n(t)$, llamados, a veces, *procesos empíricos*, con ayuda del proceso $w^\circ(t)$. Precisamente por eso podemos imaginarnos que, con grandes valores de n, tiene lugar la igualdad aproximada

$$\sqrt{n}(F_n^*(t) - F(t)) \approx w^\circ(t) \tag{4}$$

que describe la distribución de las desviaciones de $F_n(t)$ respecto a $F(t)$ (recordemos que aquí hemos considerado que $F(t) = t$, $t \in [0, 1]$.

No obstante, necesitaremos la afirmación del tipo (4) en una forma más fuerte. Examinemos, por ejemplo, la estadística $U = \sqrt{n} \sup_t (F_n^*(t) - F(t))$. Dicha afirmación hace natural la suposición de que con grandes valores de n la variable aleatoria U está distribuida aproximadamente al igual que $\sup\limits_{0 \leqslant t \leqslant 1} w^\circ(t)$. Pero de nuestra afirmación esto no se deduce de ningún modo, puesto que U no puede ser representada como función de los valores de $w^n(t) = \sqrt{n}(F_n^*(t) - F(t))$ en cualquier número finito de puntos. Por eso es mucho más fuerte la siguiente afirmación.

Designemos por $D(a, b)$ el espacio de las funciones sobre el segmento $[a, b]$, que son continuas a la izquierda (en el punto a a la derecha) y tienen sólo un número finito de saltos, y designemos por $C(a, b)$ el espacio de todas las funciones continuas sobre $[a, b]$. Es evidente que la trayectoria

$w^n(t)$ pertenece a $D(0, 1)$. Además, es sabido (véase [11], capítulo 13) que las trayectorias $w^\circ(t)$ pertenecen a $C(0, 1)$ con probabilidad 1. Para simplificar la exposición podemos suponer que todas las trayectorias $w(t)$ y, por consiguiente, $w^\circ(t)$ se encuentran en $C(0, 1)$ (véase [11]). Como $C(0, 1) \subset \subset D(0, 1)$, entonces $(D(0, 1), \sigma_D)$ — donde σ_D es el σ-álgebra de los subconjuntos de $D(0, 1)$, engendrada por conjuntos cilíndricos *) — puede ser considerado como el espacio muestral **) de los procesos w^n y w°.

Teorema 3 (teorema funcional del límite para los procesos empíricos). *Sea f la funcional que está definida sobre el espacio $D(0, 1)$ y que posee las propiedades siguientes:*

1) $f(w_n)$ y $f(w^\circ)$ *son magnitudes aleatorias (o sea, $f(y)$ realiza la aplicación medible $(D(0, 1), \sigma_D)$ en $(R, \mathfrak{B})$);*

2) *$f(y)$ es una funcional que es continua en los "puntos" del espacio $C(0, 1)$ con respecto a la métrica uniforme, o sea, $f(y_n) \to f(y)$ para $n \to \infty$ si $y \in C(0, 1)$ y $\varrho(y_n, y) = \sup\limits_{0 \leq t \leq 1} |y_n(t) - y(t)| \to 0$.*

Si estas condiciones han sido cumplidas, entonces

$$f(w^n) \Rightarrow f(w^\circ).$$

Si la funcional f es continua en la métrica uniforme en todo punto $y \in \in D(0, 1)$, la condición 1) se cumple automáticamente.

Es evidente que la funcional U, examinada anteriormente, satisface las condiciones del teorema, así que para $n \to \infty$,

$$U \Rightarrow \sup_{0 \leq t \leq 1} w^\circ(t).$$

Como en esta relación, la distribución del segundo miembro se puede hallar en forma explícita (véase, por ejemplo, [5], [58]):

$$\mathbf{P}\left(\sup_{0 \leq t \leq 1} w^\circ(t) > z\right) = e^{-2z^2},$$

obtenemos, de este modo, la expresión aproximada para la distribución de U.

El uso del teorema 3 para el cálculo de la distribución límite de otras estadísticas se examina en los párrafos siguientes.

La demostración del teorema 3 se da en el Suplemento II.

---

*) O sea, por los conjuntos que tienen la forma $\{y(t_1) \in B_1, \ldots, y(t_m) \in B_m\}$, donde $B_1, \ldots, B_m$ son los conjuntos de Borel.

**) (D_0, σ) es el espacio muestral del proceso $\xi(t)$ si en él está dada la distribución del conjunto ξ de tal modo que las trayectorias $\xi(t)$ se encuentran en D_0.

§ 7. Distribución límite para las estadísticas de primer tipo

Recordemos que llamamos estadísticas de primer tipo las estadísticas $S_n(X) = G(F_n^*)$, donde la funcional G tiene la forma $G(F) = h\left(\int g(x)\, dF(x)\right)$. Con otras palabras,

$$S_n(X) = h\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right).$$

Ya hemos visto (teorema 3.1) que si $X \in F_0$ y h es continua en el punto $a = \int g(x)\, dF_0(x)$, entonces $S_n \xrightarrow[c.s]{} h(a)$.

Teorema 1. *Si* $X \in F_0$, *h es derivable en el punto a,* $\int g^2(x)\, dF_0(x) < \infty$, *entonces*

$$\sqrt{n}(S_n(X) - h(a)) \Rightarrow h'(a)\xi,$$

donde $\xi \in \Phi_{0,\sigma^2}$, $\sigma^2 = \int (g(x) - a)^2 dF_0(x)$. Φ_{0,σ^2} *aquí significa la distribución normal con parámetros* $(0, \sigma^2)$.

Demostración. Representemos la estadística $S_n(X)$ en la forma

$$h\left(a + \frac{1}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g(x_i) - a)\right]\right),$$

donde, según el teorema central del límite (véase [11]),

$$\eta_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g(x_i) - a) \in \Phi_{0,\sigma^2},$$

$$\sigma^2 = \mathrm{M}(g(x_1) - a)^2 = \int (g(x) - a)^2 dF_0(x).$$

Nos queda hacer uso del tercer teorema de continuidad para $b_n = 1/\sqrt{n}$. ◁

A veces es más cómodo examinar las funcionales de primer tipo en la forma $G(F) = h\left(\int g(x) d(F - F_0)\right)$. Evidentemente, todo lo dicho también es válido para éstas, con la única diferencia de que a ha de considerarse igual a 0.

Citemos el análogo del teorema 1 para el caso en que la función $g = (g_1, \ldots, g_s)$ es el vector (o sea $G(F) = h\left(\int g_1(x)dF(x), \ldots, \int g_s(x)dF(x)\right)$.

Teorema 1A. *Supongamos que* $S_n(X) = G(F_n^*)$, $h(t)$ *es derivable en el*

punto $a = \int g(x)dF_0(x)$, *y que la matriz de los segundos momentos* $\sigma^2 = \|\sigma_{ij}\| = \mathbf{M}(g(x_1) - a)^T(g(x_1) - a)$ *es finita. Entonces*

$$(S_n(X) - h(a))\sqrt{n} \Rightarrow \xi(h'(a))^T = \sum_{j=1}^{s} \frac{\partial h(a)}{\partial t_j} \xi_j, \qquad (1)$$

donde $\xi = (\xi_1, \ldots, \xi_s) \in \Phi_{0,\sigma^2}$.

Si $\xi(h'(a))^T = 0$ *con probabilidad* 1, *y la matriz de segundas derivadas* $h''(t) = \left\| \frac{\partial^2}{\partial t_i \partial t_j} h(t) \right\|$ *existe en el punto a, entonces*

$$(S_n(X) - h(a))n \Rightarrow \frac{1}{2} \xi h''(a) \xi^T = \frac{1}{2} \sum_{i,j=1}^{s} \frac{\partial^2 h(t)}{\partial t_i \partial t_j} \xi_i \xi_j.$$

Para la demostración del teorema 1A conviene usar el teorema de continuidad 5.3A y el teorema central del límite multidimensional, en virtud del cual $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g(x_i) - a) \Rightarrow \xi$ (véase el suplemento V).

Completamente análogo es el teorema de la distribución límite $S_n(X)$ cuando la función h, y junto con ella también la estadística $S_n(X)$, son vectores. El lector reproducirá sin dificultad su enunciación y demostración con ayuda del teorema 5.3B.

Ejemplo 1. Supongamos que $X \in \mathbf{P}_0$ y $\mathbf{P}_0$ es tal que $\mathbf{M}x_1 = \alpha > 0$, $\mathbf{D}x_1 = d^2 < \infty$. ¿Qué representa en estas condiciones la distribución límite de la estadística $S = 1/\bar{x}\left(\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i\right)$? Aquí, las condiciones del teorema 1 están evidentemente cumplidas para $h(t) = 1/t$, $g(x) = x$, con la particularidad de que $a = \alpha$, $\sigma^2 = d^2$, $h(a) = 1/\alpha$, $h'(a) = -1/\alpha^2$. En virtud del teorema 1,

$$(S - 1/\alpha)\sqrt{n} \Rightarrow -\xi/\alpha^2, \quad \xi \in \Phi_{0,d^2},$$

así que

$$(S - 1/\alpha)\sqrt{n} \in \Phi_{0,d^2/\alpha^4}.$$

Ejemplo 2. Hallemos la distribución límite de la estadística

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

si $\mathbf{M}x_1 = \alpha$, $\mathbf{D}x_i = d^2$ y $\mathbf{M}x_1^4 < \infty$. (Ya sabemos que en virtud del primer

teorema de continuidad, $S^2 \underset{cs}{\to} d^2$). No es difícil hallar directamente la distribución límite necesaria, utilizando las representaciones

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \alpha^2) - (\bar{x} - \alpha)^2,$$

$$(S^2 - d^2)\sqrt{n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[(x_i - \alpha)^2 - d^2] - \sqrt{n}(\bar{x} - \alpha)^2.$$

No obstante, haremos uso del teorema 1A. Según los datos de este teorema debemos suponer que

$$G(F) = \int (x - \alpha)^2 dF(x) - \left(\int x\, dF(x) - \alpha\right)^2,$$

así que $g_1(x) = (x - \alpha)^2$, $g_2(x) = x$, $h(t) = t_1 - (t_2 - \alpha)^2$. Puesto que en el punto $a = (d^2, \alpha)$

$$\frac{\partial h(a)}{\partial t_1} = 1, \quad \frac{\partial h(a)}{\partial t_2} = 0,$$

entonces

$$(S^2 - d^2)\sqrt{n} \Rightarrow \xi, \quad \xi \in \Phi_{0,\upsilon^2}; \quad \upsilon^2 = \mathbf{M}(x_1 - \alpha)^4 - d^4.$$

Ejemplo 3. *Estadística* χ^2. Concluyendo este párrafo examinemos un ejemplo de estadística que puede pertenecer tanto a la del tipo I como a la del tipo II.

Examinemos las estadísticas construidas con ayuda de la funcional que tiene la forma

$$G(F) = h\left(\int g\, dF\right). \tag{2}$$

donde g es la función de variación limitada sobre el segmento $[a, b]$ tal que $F(a) = 0$, $F(b) = 1$ (a y b pueden ser infinitos). Como $\int g\, dF = g(b) - \int F\, dg$, la funcional $G(F)$ será continua en la métrica uniforme si sólo es continua la función h. Es fácil comprender que la clase destacada de características no es sino la intersección de las clases de estadísticas de los tipos I y II.

Lo mismo es válido en el caso en que g es una función de forma vectorial con componentes g_i que tienen una variación limitada.

Examinemos ahora la partición del eje real (espacio $\mathscr{X}$) en los intervalos disjuntos $\Delta_1, \ldots, \Delta_r$, y designemos $\nu_i = n\mathbf{P}_n^*(\Delta_i)$, $p_i = \mathbf{P}_0(\Delta_i)$ ($\mathbf{P}_0$ es la distribución correspondiente a F_0, así que $X \in \mathbf{P}_0$). Se llama estadística "ji-

cuadrado" $\chi^2 = \chi^2(X)$ la estadística

$$\chi^2(X) = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i} .$$

Evidentemente que esto es una estadística de tipo II, ya que ella corresponde, con una exactitud de hasta el factor n, a la funcional

$$G(F) = G_1(P) = \sum_{i=1}^{r} \frac{(\mathbf{P}(\Delta_i) - \mathbf{P}_0(\Delta_i))^2}{\mathbf{P}_0(\Delta_i)} .$$

Para representar $\chi^2(X)$ como estadística de tipo I, examinemos la funcional que tiene la forma (2)

$$G(F) = h\left(\int g\,d(F - F_0)\right)$$

con la función $h(u) = \sum_{j=1}^{r} u_j^2$ y la función vectorial g con coordenadas

$$g_j(x) = \begin{cases} 1/\sqrt{p_j} & \text{para } x \in \Delta_j, \\ 0 & \text{para } x \notin \Delta_j. \end{cases}$$

Como la función h es derivable, $\frac{\partial h(0)}{\partial u_j} = 0$, $\frac{\partial^2 h(0)}{\partial u_i \partial u_j} = 2\delta_{ij}$ (δ_{ij} es el símbolo de Kronecker), entonces, poniendo $S_n(X) = G(F_n^*)$, obtenemos

$$nS_n(X) = n \sum_{j=1}^{r} \left[\left(\frac{\nu_j}{n} - p_j\right)\frac{1}{\sqrt{p_j}}\right]^2 = \chi^2(X).$$

Para $X \in P_0$, en virtud de la segunda parte del teorema 1A,

$$\chi^2(X) \Rightarrow \sum_{j=1}^{r} \xi_j^2, \tag{3}$$

donde $\xi = (\xi_1, \ldots, \xi_r)$ es el vector normalmente distribuido (límite para $\left(\frac{\nu_1 - np_1}{\sqrt{np_1}}, \ldots, \frac{\nu_r - np_r}{\sqrt{np_r}}\right)$) con la media nula y la matriz $\sigma^2 = \|\sigma_{ij}\|$ de segundos momentos

$$\sigma_{ij} = \mathbf{M}\xi_i\xi_j = \mathbf{M}(g_i(x_1) - \sqrt{p_i})(g_j(x_1) - \sqrt{p_j})$$

(de la definición de g_j se deduce que $\mathbf{M}g_j(x_1) = \sqrt{p_j}$). Puesto que $g_i(x)g_j(x) = 0$ para $i \neq j$ y $\mathbf{P}(g_j^2(x_1) = 1/p_j) = p_j$, $\mathbf{P}(g_j^2(x_1) = 0) = 1 - p_j$, entonces

$$\sigma_{ij} = \delta_{ij} - \sqrt{p_i p_j}.$$

Aclaremos ahora qué representa la distribución del segundo miembro en (3) (o sea, la distribución límite $\chi^2(X)$).

Examinemos la transformación ortogonal en R^r con la matriz C y examinemos el vector

$$\eta = \xi C.$$

El vector η, al igual que ξ, será distribuido normalmente. En efecto, la normalidad de la magnitud ξ quiere decir que su función característica es igual a (véase [11])

$$\mathbf{M}e^{it\xi^T} = e^{-\frac{1}{2}t\sigma^2t^T},$$

donde $\sigma^2 = \|\sigma_{ij}\|$ es la matriz de segundos momentos. Pero f.c. para η

$$\mathbf{M}e^{it\eta^T} = \mathbf{M}e^{itC^T\xi^T} = e^{-\frac{1}{2}tC^T\sigma^2Ct^T}$$

tiene la misma forma y, por consiguiente, η es un vector normal, pero con la matriz de segundos momentos $d^2 = C^T\sigma^2C = \|d_{ij}\|$, así que

$$d_{ij} = \mathbf{M}\eta_i\eta_j = \sum_{k,l} c_{li}\sigma_{lk}c_{kj} = \sum_{k,l} c_{li}(\delta_{lk} - \sqrt{p_lp_k})c_{kj} =$$

$$= \sum_l c_{li}c_{lj} - \left(\sum_l c_{li}\sqrt{p_l}\right)\left(\sum_k c_{kj}\sqrt{p_k}\right). \tag{4}$$

Escojamos ahora la matriz C de modo que su primera columna tenga las coordenadas $c_{l1} = \sqrt{p_l}$ (esto corresponde a la fijación del primer vector del sistema transformado de las coordenadas y es posible, ya que $\sum_{l=1}^{r} c_{l1}^2 = \sum p_l = 1$). En este caso es evidente que el segundo sumando en (4), en virtud de la ortogonalidad de C, es igual a 1 sólo para $i = j = 1$, y es igual a 0 en el caso contrario. Esto significa que $d_{11} = \mathbf{M}\eta_1^2 = 0$, $d_{ij} = \mathbf{M}\eta_i\eta_j = \delta_{ij}$ para $i \geqslant 2$, y por consiguiente, $\eta_1 = 0$ con una probabilidad igual a 1, y las magnitudes $\eta_2, \ldots, \eta_r$ son independientes y están distribuidas normalmente con los parámetros (0, 1). A base de la ortogonalidad de C obtenemos

$$\sum_{j=1}^{r} \xi_j^2 = \sum_{j=1}^{r} \eta_j^2 = \sum_{j=2}^{r} \eta_j^2,$$

$$\chi^2(X) \Rightarrow \sum_{j=2}^{r} \eta_j^2. \tag{5}$$

En esta igualdad, la distribución del segundo miembro se llama *distribución* χ^2 ("ji-cuadrado") con $r - 1$ grados de libertad (véase [11] y también

el § 2.2). En la exposición ulterior encontraremos muchas veces esta distribución.

Una demostración más de (5) será obtenida en el párrafo siguiente. Además, (5) será demostrado en el § 3.16 con ayuda de consideraciones más generales.

Algunos otros ejemplos de uso de los teoremas 1 y 1A se dan en los capítulos posteriores.

§ 8. Distribución límite para las estadísticas de segundo tipo

Aquí nos limitaremos a examinar el caso $\mathscr{X} = R$. La funcional $G(F_n^*)$ sujeta a estudio será una magnitud aleatoria si ella realiza la aplicación medible $(D(-\infty, \infty), \sigma_D)$ en $(R, \mathfrak{B})$. Sin embargo, en lo sucesivo nos será más cómodo estudiar las funcionales que no están definidas sobre $D(-\infty, \infty)$ sino sobre $D(0, 1)$ (compárense con el § 6).

Para hacer esto apliquemos $D(-\infty, \infty)$ en $D(0, 1)$. Supongamos que la función de distribución F_0, correspondiente a la muestra, es continua y monótona, así que está definida la función inversa $F_0^{-1}(t)$ (igual a la cuantila de orden t de F_0). Nos será suficiente examinar los valores de $G(F)$ para las funciones F, cuyo portador está presente en el portador de F_0. A cada F pongámosle en correspondencia la función

$$\tilde{F}(t) = F(F_0^{-1}(t)) \equiv FF_0^{-1}(t).$$

Es evidente que $N_{\tilde{F}} \subseteq [0, 1]$, donde $N_{\tilde{F}}$ es el portador de $\tilde{F}$, así que $\tilde{F} \in D(0, 1)$ es precisamente la función de distribución. La transformación inversa de $D(0, 1)$ en $D(-\infty, \infty)$ se lleva a cabo por la igualdad

$$F(u) = \tilde{F}(F_0(u)) \equiv \tilde{F}F_0(u).$$

Pongamos ahora en correspondencia con la funcional G la funcional $\tilde{G}$ definida sobre las funciones de distribución $H \in D(0, 1)$ $(N_H \subseteq [0, 1)$ por la igualdad

$$\tilde{G}(H) = G(HF_0). \tag{1}$$

La inversión de esta fórmula tiene la forma

$$G(F) = \tilde{G}(FF_0^{-1}).$$

Estas igualdades reducen el estudio de las funcionales $G(F)$ al estudio de las funcionales $\tilde{G}(H)$ definidas en las funciones de distribución de $D(0,1)$. En virtud de estas igualdades,

$$G(F_n^*) = \tilde{G}(F_n^*F_0^{-1}) = \tilde{G}(D_n^*). \tag{2}$$

$$D_n^* = F_n^*F_0^{-1} \tag{3}$$

no es otra cosa sino la función empírica de distribución de la muestra desde la distribución uniforme sobre [0, 1]. En efecto, según el teorema 6.1, el proceso $nD_n^*(t) = nF_n^*(F_0^{-1}(t))$ tiene la misma distribución que el proceso poissoniano $\pi(F_0(F_0^{-1}(t))) = \pi(t)$, $t \in [0, 1]$ (con un parámetro $\lambda > 0$) a condición de que $\pi(1) = n$. En virtud de ese mismo teorema 6.1, esto demuestra la afirmación requerida.

Lo dicho significa que el estudio de $G(F_n^*)$ se reduce a la investigación de la funcional $\bar{G}$ de la distribución empírica que corresponde a la distribución uniforme sobre [0, 1].

Ejemplo 1. Sea $G(F) = \zeta_p$ la cuantila de orden p de la función de distribución F. Entonces $\bar{G}(H) = G(HF_0)$ será la cuantila de orden p de la función de distribución HF_0 o bien, que es lo mismo (supongamos, para simplificar, que H es continua), la solución de la ecuación $H(F_0(t)) = p$, igual a $F_0^{-1}(H^{-1}(p))$.

Esto significa que la cuantila muestral $\zeta_p^* = G(F_n^*) = \bar{G}(D_n^*)$ (véanse (2) y (3)) de la muestra $X \in F_0$ no es otra cosa sino el valor de la función F_0^{-1} de la cuantila muestral $\eta_p^* = (D_n^*)^{-1}(p)$ de orden p de la muestra Y de la distribución uniforme.

Por lo tanto, si logramos hallar la distribución límite de η_p^*, entonces la distribución límite de ζ_p^* podrá ser obtenida con ayuda de los teoremas de continuidad.

Ejemplo 2. Examinemos la funcional $G(F) = \sup_{-\infty<t<\infty} |F(t) - F_0(t)|$. En este caso

$$\bar{G}(H) = G(HF_0) = \sup_{-\infty<t<\infty} |H(F_0(t)) - F_0(t)| = \sup_{u\in[0,1]} |H(u) - u|,$$

así que

$$G(F_n^*) = G(D_n^*) = \sup_{u\in[0,1]} |D_n^*(u) - u|,$$

y en correspondencia con el contenido del § 6, la distribución de la estadística $G(F_n^*)$ no dependerá de F_0 si F_0 es continua. En este sentido la estadística $G(F_n^*)$ puede llamarse invariante respecto a la distribución uniforme de la muestra.

Ejemplo 3. La funcional

$$G(F) = \int_{-\infty}^{\infty} |F(t) - F_0(t)|^k dF_0(t)$$

también engendra la estadística $G(F_n^*)$, invariante respecto a F_0, ya que

$$\bar{G}(H) = \int_0^1 |H(u) - u|^k du, \qquad G(F_n^*) = \int_0^1 |D_n^*(u) - u|^k du.$$

Ejemplo 4. Examinemos la funcional

$$G(F) = \sum_{j=1}^{r} \frac{(\Delta_j F - \Delta_j F_0)^2}{\Delta_j F_0},$$

donde $\Delta_j F$ son los incrementos de la función F sobre los intervalos $\Delta_j = [t_j, t_{j+1})$ que forman la partición de una recta real. Evidentemente que $nG(F_n^*)$ no es otra cosa sino la estadística χ^2 examinada en el ejemplo 7.3 en calidad de estadística de tipo I.

Tenemos

$$\tilde{G}(H) = G(HF_0) = \sum_{j=1}^{r} \frac{(\Delta_j HF_0 - \Delta_j F_0)^2}{\Delta_j F_0},$$

donde

$$\Delta_j HF_0 = H(F_0(t_{j+1})) - H(F_0(t_j)) = \delta_j H,$$

$\delta_j H$ son los incrementos de H sobre los intervalos $\delta_j = [\tau_j, \tau_{j+1})$, $\tau_j = F_0(t_j)$. Así, pues, designando con esa misma letra δ_j la longitud del intervalo δ_j, obtenemos

$$G(F_n^*) = \tilde{G}(F_n^* F_0) = \tilde{G}(D_n^*) = \sum_{j=1}^{r} (\delta_j D_n^* - \delta_j)^2/\delta_j.$$

Aquí el segundo miembro es la estadística $n^{-1}\chi^2$ para la muestra Y de la distribución uniforme con partición $\{\delta_j\}$. Esto significa, en particular, que en el ejemplo 3 del párrafo precedente pudiéramos limitarnos a examinar la distribución uniforme F_0, aunque la estadística χ^2 por sí misma no es invariante con respecto a F_0.

Ahora bien, podemos, sin limitar la generalidad, suponer que la funcional $G(F)$ se da sobre $D(0, 1)$ y $F_0(t) = t$, $t \in [0, 1]$. El paso a las funcionales "iniciales" se realiza mediante las fórmulas (1) y (2) y será ilustrado con otros ejemplos.

Con el fin de encontrar la distribución límite para las funcionales de segundo tipo $G(F_n^*)$ es necesario, al igual que en el apartado precedente, imponer a las funcionales ciertas condiciones de suavidad.

Pongamos para abreviar, $\|x\| = \sup_{0 \leq t \leq 1} |x(t)|$.

Definición 1. La funcional $G(F)$ se llama *continuamente derivable de orden k en el punto* F_0 si existe la funcional $g(F_0, v)$ que para cualquier función $v \in C(0, 1)$ y cualquier sucesión $v_h \in D(0, 1)$ es tal que $\|v_h - v\| \to 0$ cuando $h \to 0$ satisface las relaciones

$$\frac{G(F_0 + hv_h) - G(F_0)}{h^k} \to g(F_0, v), \tag{4}$$

$$g(F_0, v_h) \to g(F_0, v).$$

La última relación significa, evidentemente, la continuidad en la métrica uniforme en los puntos de $C(0, 1)$ de la funcional $g(F_0, v)$ que se puede llamar *derivada de orden k de G en la dirección de v*.

Observación 1. Recordemos que aquí, en cualquier parte, por F_0 se puede entender la distribución uniforme sobre $[0, 1]$.

Mostremos que en el ejemplo 1, la funcional $G(F) = F^{-1}(p)$ de la distribución F sobre $[0, 1]$ es continuamente derivable en el "punto" $F_0(t) = t$, $t \in [0, 1]$.

En efecto, por definición,

$$G(F_0 + hv_h) = \text{máx}\ \{t\colon F_0(t) + hv_h(t) \leqslant p\}.$$

Como esta funcional es continua en la métrica uniforme en el punto F_0, podemos poner $G(F_0 + hv_h) = p + \delta$, donde $\delta = \delta(h) \to 0$ para $h \to 0$. Luego, de la relación $\|v_h - v\| \to 0$, donde $v \in C(0, 1)$, se deduce $|v_h(p + \delta) - v_h(p)| = r(h) \to 0$ cuando $h \to 0$. Como $F_0(p + \delta) = p + \delta$, para $t = G(F_0 + pv_h) = p + \delta$ obtenemos

$$F_0(t) + hv_h(t) = p + \delta + hv_h(p + \delta) = p + \delta + h(v_h(p) + \tau r(h)) \leqslant p,$$

donde $|\tau| \leqslant 1$. La igualdad inversa análoga se puede escribir valiéndose del hecho de que $F_0(t + 0) + hv_h(t + 0) \geqslant p$. De aquí se deduce que $\delta = -h(v_h(p) + \tau_1 r(h))$, $|\tau_1| \leqslant 1$, así que

$$\frac{G(F_0 + hv) - G(F_0)}{h} = \frac{\delta}{h} \to -v(p).$$

Ahora bien, la derivada $g(F_0, v)$ en este ejemplo es igual a

$$g(F_0, v) = -v(p). \ \lhd \tag{5}$$

Es evidente que en el ejemplo 2, la funcional $G(F) = \sup\limits_{t \in [0, 1]} |F(t) - F_0(t)|$ es también continuamente derivable en toda dirección, ya que $G(F_0) = 0$,

$$g(F_0, v) = \frac{G(F_0 + hv)}{h} = \sup_{t \in [0,1]} |v(t)|.$$

En el ejemplo 3, la funcional $G(F) = \int\limits_0^1 |F(t) - F_0(t)|^k dR(t)$ para cualquier función de variación limitada $R(t)$ es continuamente derivable (de orden k) en toda dirección, ya que

$$g(F_0, v) = \frac{G(F_0 + hv)}{h^k} = \int\limits_0^1 |v(t)|^k dR(t).$$

La afirmación análoga es válida respecto al ejemplo 4 sobre la funcional

$$G(F) = \sum_{j=1}^{r} \frac{(\Delta_j F - \Delta_j F_0)^2}{\Delta_j F_0}$$

la cual será continuamente derivable de segundo orden, puesto que para ella

$$g(F_0, v) = \frac{G(F_0 + hv)}{h^2} = \sum_{j=1}^{r} \frac{(\Delta_j v)^2}{\Delta_j F_0}.$$

En los ejemplos 2 — 4, la generalización de las funcionales son las funcionales de forma $G(F) = G_1(F - F_0)$, donde la funcional G_1 es homogénea en el sentido de que $G_1(hv) = h^k G(v)$. Es evidente que todas estas funcionales serán derivables.

Enunciemos ahora el teorema principal de las funcionales de segundo tipo. Sea, como antes, $F_0(t) \equiv t$, $t \in [0, 1]$.

Teorema 1. *Si $X \in F_0$ y la funcional $G(F)$ es derivable (de orden k) en sentido de la definición* 1, *entonces*

$$[G(F_n^*) - G(F_0)]n^{k/2} \Rightarrow g(F_0, w^\circ),$$

donde w° es el puente browniano.

Demostración. Es sabido (véase, por ejemplo, [5]) que los compactos en el espacio métrico de las funciones continuas $C(0, 1)$ con métrica uniforme, se describen del modo siguiente. A cada función $\varphi(\Delta) > 0$, $\varphi(\Delta) \to 0$ para $\Delta \to 0$, y al número $N > 0$ le corresponde el compacto

$$K = K(\varphi, N) = \{y \in C(0, 1): \omega_\Delta(y) \leqslant \varphi(\Delta), |y(0)| \leqslant N\},$$

donde $\omega_\Delta(y)$ es el módulo de continuidad y:

$$\omega_\Delta(y) = \sup_{|t-u|\leqslant\Delta} |y(t) - y(u)|.$$

Designemos por K_h el conjunto

$$K_h = \{y \in D(0, 1): \omega_\Delta(y) \leqslant \varphi(\Delta) \text{ para todos } \Delta \geqslant h: |y(0)| \leqslant N\}.$$

Los conjuntos K_h podrían llamarse "precompactos" (este término se utiliza en el análisis funcional en otro sentido) engendrados por el compacto K. Está claro que $K_{h_1} \subset K_{h_2}$ para $h_1 \leqslant h_2$, $\bigcap_{n=1}^{\infty} K_{1/n} = K$ y que $K_h \subset (K)^{\varphi(h)}$, donde $(K)^\varepsilon$ es el ε-entorno del conjunto K.

Mostremos ahora que para $\delta > 0$ dado existe el compacto K (y, por lo tanto, la familia de los precompactos K_h que le corresponden) y la suce-

sión $h_n \to 0$ para $n \to \infty$ tales que

$$\lim_{n\to\infty} \sup \mathbf{P}(w^n \notin K_{h_n}) \leqslant \delta. \tag{6}$$

En efecto, según el teorema 6.3, para toda funcional f que sea continua en la métrica uniforme se cumple $f(w^n) \Rightarrow f(w^\circ)$, donde $w^n(t) = \sqrt{n}(F_n^*(t) - t)$, $0 \leqslant t \leqslant 1$. Como $\omega_\Delta(y)$ es tal funcional, entonces $\omega_\Delta(w^n) \Rightarrow \omega_\Delta(w^\circ)$. Pero $\omega_\Delta(w^\circ) \underset{c.s}{\to} 0$ para $\Delta \to 0$, ya que las trayectorias de w° son casi seguramente continuas. Por consiguiente, para ε y δ dados, siendo Δ suficientemente pequeño,

$$\mathbf{P}(\omega_\Delta(w^\circ) > \varepsilon) \leqslant \delta.$$

Considerando, sin limitar la generalidad, el número ε como punto de continuidad de la distribución $\omega_\Delta(w^\circ)$, obtenemos

$$\lim_{n\to\infty} \sup \mathbf{P}(\omega_\Delta(w^n) > \varepsilon) \leqslant \delta.$$

Sea ahora $\varepsilon_k \downarrow 0$ cierta sucesión, y los números $\Delta_k \downarrow 0$ son tales que

$$\lim_{n\to\infty} \sup \mathbf{P}(\omega_{\Delta_k}(w^n) > \varepsilon_k) \leqslant \delta/2^{k+1}.$$

Formemos la función $\varphi(\Delta) = \varepsilon_k$ para $\Delta \in [\Delta_{k+1}, \Delta_k)$. Es evidente que $\varphi(\Delta) \to 0$ para $\Delta \to 0$, y podemos examinar los precompactos K_h construidos según la función φ. Entonces para todo $k < \infty$,

$$\lim_{n\to\infty} \sup \mathbf{P}(w^n \notin K_{\Delta_k}) \leqslant \lim_{n\to\infty} \sup \sum_{j=1}^{k+1} \mathbf{P}(\omega_{\Delta_j}(w^n) > \varepsilon_j) \leqslant$$

$$\leqslant \sum_{j=1}^{k+1} \lim_{n\to\infty} \sup \mathbf{P}(\omega_{\Delta_j}(w^n) > \varepsilon_j) \leqslant \delta/2$$

(para $k = \infty$ esta desigualdad puede ser injusta). La relación obtenida quiere decir que para cada δ existe la sucesión $h_n \to 0$ cuando $n \to \infty$ es tal que se cumple (6). Examinemos ahora la magnitud

$$[G(F_n^*) - G(F_0)]n^{k/2} = g(F_0, w^n) + H_n(w^n),$$

donde $H_n(x) = [G(F_0 + x/\sqrt{n}) - G(F_0)]n^{k/2} - g(F_0, x)$. Puesto que, en virtud del teorema 6.3 y la definición 1, $g(F_0, w^n) \Rightarrow g(F_0, w^\circ)$, basta con que nos cercioremos de que

$$H_n(w^n) \underset{P}{\to} 0. \tag{7}$$

Nótese que para todo compacto $K \subset C(0, 1)$ y para toda sucesión $h_n \to 0$ cuando $n \to \infty$,

$$\sup_{\substack{x \in D(0,1) \\ x \in (K)^{h_n}}} |H_n(x)| \to 0.$$

Admitiendo lo contrario, llegaremos a la existencia de una sucesión $x_n \in D(0, 1)$ tal que $|x_n - x| \to 0$, $x \in C(0, 1)$, $\varlimsup_{n\to\infty} |H_n(x_n)| > 0$, lo cual contradice la derivabilidad de G.

A base de (6) y (8) obtenemos

$$\mathbf{P}(|H_n(w^n)| > \varepsilon) \leqslant \mathbf{P}(|H_n(w^n)| > \varepsilon, \quad w^n \in K_{h_n}) + \mathbf{P}(w^n \notin K_{h_n}),$$

$$\varlimsup_{n\to\infty} \mathbf{P}(|H_n(w^n)| > \varepsilon) \leqslant \delta.$$

Como δ es arbitrario, la relación (7) y junto con ella la afirmación del teorema quedan demostradas. ◅

Volvamos a examinar los ejemplos.

Sea η_p^* la cuantila muestral de orden p para la muestra Y de la distribución uniforme sobre [0, 1]. Entonces, de (5) y del teorema 1 obtenemos que

$$(\eta_p^* - p)\sqrt{n} \Rightarrow -w^\circ(p) \underset{d}{=} w^\circ(p).$$

Hemos determinado, además, que en el caso general, cuando F_0 es una función continua arbitraria de distribución, es válida la igualdad

$$\zeta_p^* = F_0^{-1}(\eta_p^*).$$

Si ahora utilizamos el tercer teorema de continuidad, obtendremos:

Corolario 1. *Si $X_n \in F_0$, F_0 es continuamente derivable en el punto ζ_p, $f(\zeta_p) = F_0'(\zeta_p) > 0$, entonces*

$$(\zeta_p^* - \zeta_p)\sqrt{n} \Rightarrow w^\circ(p)/f(\zeta_p).$$

Para la demostración sólo es necesario señalar que las condiciones del corolario 1 significan la derivabilidad continua de F_0^{-1} en el punto p,

$$(F_0^{-1}(p))' = \frac{1}{F_0'(F_0^{-1}(p))} = \frac{1}{f(\zeta_p)}.$$

Como $\mathbf{M}w^\circ(p) = 0$, $\mathbf{D}w^\circ(p) = \mathbf{M}(w(p) - pw(1))^2 = \mathbf{M}(w(p)(1-p) + p(w(1) - w(p)))^2 = p(1-p)^2 + p^2(1-p) = p(1-p)$, la afirmación del corolario 1 también puede escribirse en la forma

$$(\zeta_p^* - \zeta_p)\sqrt{n} \in \Phi_{0,\sigma^2}, \quad \sigma^2 = p(1-p)/f^2(\zeta_p). \ ◅$$

En el ejemplo 2 derivamos la funcional $G(F) = \sup_{0\leqslant t\leqslant 1} |F(t) - F_0(t)|$ y, por lo tanto, según el teorema 1,

$$G(F_n^*)\sqrt{n} \Rightarrow \sup_{0\leqslant t\leqslant 1} |w^\circ(t)|.$$

Hemos hallado la distribución $\eta = \sup_{0 \leq t \leq 1} |w^{\circ}(t)|$ en forma explícita ([58]):

$$\mathbf{P}(\eta > z) = K(z) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2z^2}.$$

La función $K(z)$ se llama *función de Kolmogórov.*

Hemos visto que en el caso general, cuando F_0 es una función continua arbitraria de distribución, la distribución de la estadística

$$D(x) = \sup_t |F_n^*(t) - F_0(t)|$$

queda igual que para el caso $F_0(t) = t$, $t \in [0, 1]$. De este modo hemos obtenido:

Corolario 2 (teorema de Kolmogórov). *Si $X \in F_0$, F_0 es continua, entonces*

$$\sqrt{n}\, D(X) \in K.$$

Esto significa que la desviación máxima $D(X)$ de la función $F_n^*(t)$ de $F_0(t)$ tiene el orden $1/\sqrt{n}$ y puede representarse, aproximadamente, en la forma de $D(X) \approx \eta/\sqrt{n}$.

En el ejemplo 3 hemos visto que otra estadística (la cual a menudo se designa por ω^2)

$$\omega^2 = \int_{-\infty}^{\infty} (F_n^*(t) - F_0(t))^2 dF_0(t)$$

también es invariante respecto a F_0. Del teorema 1 se deduce:

Corolario 3. *Si $X \in F_0$, F_0 es continua, entonces*

$$n\omega^2 \Rightarrow \int_0^1 [w^{\circ}(t)]^2 dt.$$

La distribución $\int_0^1 [w^{\circ}(t)]^2 dt$ también fue hallada en forma explícita y, junto con la distribución $K(z)$, está tabulada. Con arreglo al ejemplo 4, el teorema 1 nos da:

Corolario 4. *Si $X \in F_0$, F_0 es continua, entonces*

$$\chi^2 \Rightarrow \sum_{j=1}^{r} (\delta_j w^{\circ})^2/\delta_j,$$

donde δ_j, $j = 1, 2, \ldots, r$, forman la partición del segmento $[0, 1]$ *y están definidos en el ejemplo* 4.

Si suponemos que $\xi = (\xi_1, \ldots, \xi_r)$, $\xi_j = \delta_j w^{\circ}/\sqrt{\delta_j}$ utilizando el hecho

de que $\delta_j w^\circ = \delta_j w - w(1)\delta_j$, donde w es el proceso wieneriano estándar, obtenemos

$$\chi^2 \Rightarrow \sum_{j=1}^{r} \xi_j^2, \ \xi \in \Phi_{0,\sigma^2}.$$

Aquí $\sigma^2 = |\sigma_{ij}|$ es la misma matriz que en el ejemplo 7.3, puesto que

$$\delta_j w^\circ = \delta_j w - \left(\sum_k \delta_k w\right)\delta_j = \sum_{k=1}^{r} a_{kj}\delta_k w,$$

$$a_{kj} = \delta_{kj} - \delta_j, \ \mathbf{M}(\delta_k w)(\delta_l w) = \delta_{kl}\delta_k$$

(δ_{kl} es el símbolo de Kronecker),

$$\sigma_{ij} = \frac{\mathbf{M}(\delta_i w^\circ)(\delta_j w^\circ)}{\sqrt{\delta_i\delta_j}} = \frac{1}{\sqrt{\delta_i\delta_j}} \sum_{k=1}^{r} a_{ki}a_{kj}\delta_k =$$

$$= \frac{1}{\sqrt{\delta_i\delta_j}}(\delta_{ij}\delta_i - \delta_i\delta_j) = \delta_{ij} - \sqrt{\delta_i\delta_j}.$$

Repitiendo los razonamientos del ejemplo 7.3 obtenemos que $\sum_{j=1}^{r} \xi_j^2$ tiene una distribución χ^2 con $r - 1$ grados de libertad.

Concluyendo este párrafo debemos señalar que no todas las estadísticas que representen interés pueden ser clasificadas como estadísticas de los tipos I ó II. Basta con examinar, por ejemplo, la estadística $S(X) = \sum_{i=1}^{n-1} x_i x_{i+1}$ o las estadísticas S relacionadas con las funcionales $G_n(F)$, donde las funcionales G_n dependen "considerablemente" de n (no sólo por la muestra), tales como, digamos, el término máximo de la serie variacional $S(X) = x_{(n)} = \zeta^*_{1-1/n}$ y otras.

§ 9*. Objeciones acerca de las estadísticas no paramétricas

Hay una propiedad respecto a la cual la estadística ζ^*_p en el ejemplo 8.1 se distingue considerablemente de las citadas en los ejemplos 8.2 — 8.4. Esta propiedad consiste en que la distribución límite de las estadísticas en los ejemplos 8.2 — 8.4 (véanse los corolarios 8.2 — 8.4) de ningún modo está relacionada con la función de distribución F_0, lo cual no se puede decir de la estadística ζ^*_p (compárese con el corolario 8.1).

Definición 1. La estadística $S(X)$ se llama *asintóticamente no paramétrica* si $S(X) \Leftarrow Q$ cuando $n \to \infty$, y Q no depende de la distribución de X, o sea, no depende de F_0 si $X \in F_0$.

Cabe señalar que la propia función S en este caso puede depender de F_0. El término "no paramétrica" no es por sí mismo del todo acertado, no obstante, adquirió gran divulgación (está justificado en el caso en que F_0 pertenece a cierta familia paramétrica — entonces la distribución $\mathbf{Q}$ no depende del parámetro y desde este punto de vista no es paramétrica). A veces se utiliza otro término: "libre de la distribución".

En los §§ 6—8 hemos visto que las estadísticas $\sqrt{n}\, U(X)$, $\sqrt{n}\, D(X)$, $n\omega^2(X)$, $\chi^2(X)$ son asintóticamente no paramétricas.

También debemos indicar que el teorema 6.1 da la posibilidad de introducir un concepto más estrecho. En dicho teorema se ha establecido que $nF_n^*(t)$ está igual distribuida que $\eta(F_0(t))$, donde $\eta(u)$ es el proceso poissoniano convencional con un parámetro arbitrario $\lambda > 0$ a condición de que $\eta(1) = n$ (véase el § 6), o sea, dicho proceso no depende de F_0. Ahora bien, si la estadística S está construida como la funcional $G(F_n^*)$ (o $G(F_n^* - F_0)$), que es invariante respecto a la sustitución del "tiempo" t en el argumento, la distribución de S no dependerá de F_0. Por ejemplo,

$$D = \sup_t |F_n^*(t) - F_0(t)| = \frac{1}{n} \sup_t |\eta(F_0(t)) - nF_0(t)| =$$

$$= \frac{1}{n} \sup_{u \in [0,1]} |\eta(u) - un|. \qquad (1)$$

Lo dicho hace posible:

Definición 2. La estadística $S(X)$ se llama *no paramétrica* si su distribución no depende de $F_0 (X \in F_0)$.

Las relaciones (1) significan que la estadística D no es paramétrica.

También hemos señalado (véase el corolario 8.3) que la estadística ω^2, al igual que D, no depende de F_0 y, por lo tanto, tampoco es paramétrica.

La estadística χ^2, siendo asintóticamente no paramétrica no poseerá la propiedad de carácter no paramétrico. De esto es fácil convencerse directamente en un ejemplo, poniendo $r = 2$, $n = 1$.

Obtenemos otros ejemplos de las estadísticas no paramétricas si examinamos los valores de $F_n^*(\zeta_p)$, donde ζ_p es la cuantila de orden p, así que $nF_n^*(\zeta_p) \underset{d}{=} \eta(p)$ (véase el § 6). El número r_j de elementos de la muestra X, menores que x_j — la llamada estadística de rango — también será una estadística no paramétrica.

Los conceptos de estadísticas no paramétrica y asintóticamente no paramétrica son muy útiles en la teoría de la verificación de las hipótesis estadísticas (véase el capítulo 3), ya que la distribución de estas estadísticas, la cual es necesaria para la construcción de los criterios, es suficiente calcularla sólo una vez (por jemplo, para la distribución uniforme de F_0) y será útil para cualesquiera otras distribuciones de la muestra.

§ 10*. Distribuciones empíricas suavizadas. Densidades empíricas

En el § 2 a cada muestra X la hemos puesto en correspondencia con la distribución $\mathbf{P}_n^*$ que hemos llamado empírica y la cual no es más que la suma de n distribuciones atómicas concentradas en los puntos $x_1, ..., x_n$. Esta distribución posee varias propiedades magníficas descritas en los párrafos precedentes. Sin embargo, la definición de $\mathbf{P}_n^*$, utilizada por nosotros, no es la única posible ni mucho menos, y en varios casos no es la más natural. También existen otros puntos de vista en cuanto a la definición de $\mathbf{P}_n^*$, según los cuales las propiedades útiles (estudiadas anteriormente) de las distribuciones empíricas no sólo se conservan por completo, sino que son completadas por varias nuevas.

Aquí nos limitaremos a examinar la cuestión relacionada con la naturaleza de las distribuciones que situamos en los puntos x_i. En la definición de $\mathbf{P}_n^*$ que hemos utilizado, se trataba de las distribuciones degeneradas $\mathbf{I}_{x_i}(B)$, así que

$$\mathbf{P}_n^*(B) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_{x_i}(B). \tag{1}$$

En este caso la distribución empírica es singular con respecto a la medida de Lebesgue y, por lo tanto, no tiene densidad. Esto puede resultar incómodo en los casos cuando sabemos de antemano que la distribución inicial $\mathbf{P}$ tiene densidad. Con arreglo a esta condición sería conveniente tener una distribución empírica suave $\mathbf{P}_n^*$ para la cual, junto con la convergencia $\mathbf{P}_n^* \to \mathbf{P}$, desde todos los puntos de vista establecidos anteriormente también tenga lugar la convergencia de las densidades $f_n^* \to f$, donde f_n^* y f son las densidades correspondientes a $\mathbf{P}_n^*$ y $\mathbf{P}$.

No es difícil obtener esto del modo siguiente. Sea $\mathbf{Q}$ cierta distribución *que tiene densidad*. Pongamos

$$\mathbf{P}_n^{**}(B) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{Q}\left(\frac{B - x_i}{h_n}\right), \tag{2}$$

donde $\dfrac{B - x}{h}$ es el conjunto de puntos $y \in \mathscr{X}$ para los cuales $x + yh \in B$; $h_n \to 0$ cuando $n \to \infty$.

Es evidente que $\mathbf{P}_n^{**}(B)$ no es otra cosa sino la "suma media" de las distribuciones $\mathbf{Q}$ contraídas hasta las dimensiones h_n y "situadas" en los puntos x_i. La definición (2) generaliza (1). La fórmula (1) se obtiene de (2) si se pone $\mathbf{Q} = \mathbf{I}_0$, ya que $\mathbf{I}_{x_i}(B) = \mathbf{I}_0(B - x_i) = \mathbf{I}_0\left(\dfrac{B - x_i}{h_n}\right)$ para cualquier sucesión $\{h_n\}$.

Señalemos las siguientes propiedades de la distribución $\mathbf{P}_n^{**}$ que llamaremos *distribución empírica suavizada.*

1. La distribución $\mathbf{P}_n^{**}$ es la convolución de las distribuciones $\mathbf{P}_n^*$ y $\mathbf{Q}(B/h_n)$, y

$$\mathbf{P}_n(B) = \mathbf{MP}_n^{**}(B) = \int \mathbf{Q}\left(\frac{B-y}{h_n}\right)\mathbf{P}(dy)$$

es la convolución de las distribuciones $\mathbf{P}$ y $\mathbf{Q}(B/h_n)$. Con otras palabras, $\mathbf{P}_n(B)$ es la distribución de la variable aleatoria $\xi + h_n\eta$, donde $\xi \in \mathbf{P}$, $\eta \in \mathbf{Q}$. De los teoremas de continuidad se deduce que para $h_n \to 0$,

$$\mathbf{P}_n \Rightarrow \mathbf{P}. \tag{3}$$

Recordemos que para la distribución $\mathbf{P}_n^*$ hemos tenido la igualdad exacta

$$\mathbf{MP}_n^* = \mathbf{P}.$$

2. Si la distribución $\mathbf{P}$ es absolutamente continua con respecto a la medida de Lebesgue, la distribución $\mathbf{P}_n^{**}$ satisfará los teoremas análogos al de Glivenko — Cantelli. En efecto, en este caso la convergencia (3) significará la convergencia uniforme de las distribuciones sobre todos los intervalos. Para simplificar la exposición nos limitaremos a un caso unidimensional, supongamos que ($F_n^{**}(x)$, $F_n(x)$ y $Q(x)$ designan las funciones de distribución correspondientes a $\mathbf{P}_n^{**}$, $\mathbf{P}_n$ y $\mathbf{Q}$)

$$F_n^{**}(x) - F(x) = \int \mathbf{Q}\left(\frac{x-y}{h_n}\right)dF_n^*(y) - F(x) =$$

$$= -\int F_n^*(y)d_yQ\left(\frac{x-y}{h_n}\right) - F(x) =$$

$$= F_n(x) - F(x) - \int (F_n^*(y) - F(y))d_yQ\left(\frac{x-y}{h_n}\right).$$

Aquí, como ya hemos señalado, la diferencia $F_n(x) - F(x) \to 0$ es uniforme en x, y la integral presente en el segundo miembro no excede $\sup_y |F_n^*(y) - F(y)| \xrightarrow{c.s.} 0$.

3. La ventaja de $\mathbf{P}_n^{**}$ en comparación con $\mathbf{P}_n^*$, por cuya razón hemos introducido la primera distribución, consiste en que *esta distribución tiene la densidad.*

$$f_n^*(x) = \frac{1}{nh_n}\sum_{i=1}^{n} q\left(\frac{x-x_i}{h_n}\right) = \frac{1}{h_n}\int q\left(\frac{x-y}{h_n}\right) dF_n^*(y) \tag{4}$$

($q(x)$ es la densidad de la distribución $\mathbf{Q}$) que para cada x, cuando $n \to \infty$ y $h_n \to 0$, se aproxima a la densidad $f(x)$ de la distribución $\mathbf{P}$.

Antes de demostrar la afirmación correspondiente, cabe señalar que para la obtención de buenos resultados acerca de la aproximación de $f_n^*(x)$ a $f(x)$, conviene utilizar las densidades limitadas suaves q. Al elegir, digamos, q indefinidas, la estimación $f_n^*(x)$ de la densidad suave $f(x)$ empeorará premeditadamente. Como la elección de q está en nuestras manos, podemos considerar que, por lo menos, queda cumplida la condición

$$d^2 = \int q^2(t)dt < \infty. \tag{5}$$

Teorema 1. *Si q satisface la condición (5), $f(x)$ es continua y limitada, $h_n \to 0$ para $n \to \infty$ de modo que $nh_n \to \infty$,* entonces

$$f_n^*(x) = f_n(x) + \zeta_n(x)/\sqrt{nh_n}, \tag{6}$$

donde $f_n(x)$ es la función no aleatoria

$$f_n(x) = \mathbf{M}f_n^*(x) = \mathbf{M}h_n^{-1}q\left(\frac{x - x_l}{h_n}\right) = \frac{1}{h_n}\int q\left(\frac{x-t}{h_n}\right)f(t)dt =$$

$$= \int q(z)f(x - zh_n)dz \to f(x) \tag{7}$$

para $h_n \to 0$. Las variables aleatorias $\zeta_n(x)$ son normales asintóticamente, $\zeta_n(x) \in \Phi_{0,\sigma^2(x)}$, $\sigma^2(x) = f(x)d^2$.

Demostración. La suma en (4) es la suma de variables aleatorias independientes e igualmente distribuidas en el esquema de series, con la particularidad de que $f_n(x) = \mathbf{M}f_n^*(x)$ está representada en (7). Pongamos

$$\xi_{k,n} = \frac{1}{\sqrt{nh_n}}\left[q\left(\frac{x - x_k}{h_n}\right) - h_n f_n(x)\right].$$

Entonces

$$f_n^*(x) - f_n(x) = \frac{1}{\sqrt{nh_n}}\sum_{k=1}^{n}\xi_{k,n}, \quad \mathbf{M}\xi_{k,n} = 0$$

$$\mathbf{M}\xi_{k,n}^2 = \frac{1}{n}\left[\mathbf{M}\frac{1}{h_n}q^2\left(\frac{x - x_k}{h_n}\right) - h_n f_n^2(x)\right],$$

$$\mathbf{M}\frac{1}{h_n}q^2\left(\frac{x - x_k}{h_n}\right) = \frac{1}{h_n}\int q^2\left(\frac{x-t}{h_n}\right)f(t)dt =$$

$$= \int q^2(z)f(x - zh_n)dz \to f(x)\int q^2(z)dz = f(x)d^2. \tag{8}$$

Ahora bien, $\mathbf{M}\xi_{k,n}^2 \sim f(x)d^2/n$ si $f(x) > 0$. La condición de Lindeberg tiene en nuestro caso la forma

$$n\mathbf{M}(\xi_{1,n}^2;\ |\xi_{1,n}| > \varepsilon) \to 0 \tag{9}$$

para $n \to \infty$ y para cualquier $\varepsilon > 0$. Como $h_n f_n^2(x) \to 0$, $n\xi_{1,n}^2 \leqslant 2(q^2((x - x_1)/h_n) + h_n f_n^2(x))$, entonces para cumplir (9) es suficiente que

$$\mathrm{M}\left(\frac{1}{h_n} q^2\left(\frac{x - x_1}{h_n}\right)\right); q\left(\frac{x - x_1}{h_n}\right) > \varepsilon\sqrt{nh_n} \to 0.$$

Esta relación tiene lugar, ya que su primer miembro es igual a (compárese con (8))

$$\int_{q(z) > \varepsilon\sqrt{nh_n}} q^2(z) f(x - zh_n)dz \leqslant c \int_{q(z) > \varepsilon\sqrt{nh_n}} q^2(z)dz \to 0.$$

Ahora bien, a la variable aleatoria $\zeta_n(x) = \sum_{k=1}^{n} \xi_{k,n}$ es aplicable el teorema central del límite. Esto demuestra el teorema 1. ◁

En el problema sujeto a examen surge naturalmente la cuestión acerca de la elección óptima de h_n y de la función $q(t)$. Sin embargo, su solución depende de las propiedades de suavidad de $f(x)$. En efecto, supongamos, por ejemplo, que $f(x)$ es positiva solamente en el intervalo finito y que es dos veces continuamente derivable con el valor fijo $\varphi = \int (f''(x))^2 dx$. Supongamos también que $\int zq(z)dz = 0$ (esto es siempre así para las $q(z)$ simétricas) y que $D^2 = \int z^2 q(z)dz < \infty$. Entonces

$$f_n(x) = \int q(z) f(x - zh_n)dz =$$

$$= \int q(z) \left[f(x) - zh_n f'(x) + \frac{z^2 h_n^2}{2} f''(x) + o(z^2 h_n^2)\right] dz =$$

$$= f(x) + \frac{h_n^2 f''(x)}{2} \int z^2 q(z)dz + o(h_n^2).$$

Vemos que

$$f_n^*(x) - f(x) = \frac{D^2 h_n^2 f''(x)}{2} + \frac{\zeta_n(x)}{\sqrt{nh_n}} + o(h_n^2),$$

$$\mathrm{M}[f_n^*(x) - f(x)]^2 = \left(\frac{D^2 h_n^2 f''(x)}{2}\right)^2 + \frac{d^2 f(x)}{nh_n} + o(h_n^4). \quad (10)$$

La minimización de esta expresión en h_n y q dará, en virtud de la normalidad asintótica de $\zeta_n(x)$, la "dispersión" mínima posible de $f_n^*(x)$ alrededor del valor de $f(x)$. No obstante, en este caso los valores minimizantes de h_n y q dependerán de x mediante los valores desconocidos de $f(x)$ y $f''(x)$. Para evitar este efecto y obtener la optimalidad "por término medio" es

natural examinar la integral

$$\int \mathbf{M}[f_n^*(x) - f(x)]^2 dx \tag{11}$$

cuya parte principal será igual a $\left(\frac{D^2 h_n^2}{2}\right)^2 \varphi + \frac{d^2}{nh_n}$ (esto se obtiene si en (10) se retira $o(h_n^4)$).

El mínimo de esta expresión se alcanza cuando $h_n = \left(\frac{d^2}{nD^4\varphi}\right)^{1/5}$. Con tal elección de h_n, la integral (11) será igual a

$$\frac{5}{4}\varphi^{1/5}(Dd^2)^{4/5} n^{-4/5} + o(n^{-4/5}), \tag{12}$$

$$-f_n^*(x) - f(x) = \left(\frac{Dd^2}{n\varphi}\right)^{2/5}\left(\frac{f''(x)}{2} + f(x)\sqrt{\varphi}\xi_n + o(n^{-2/5})\right),$$

$$\xi_n \in \Phi_{0,1}.$$

Ahora bien, aquí la velocidad de convergencia constituye sólo $n^{-2/5}$ a diferencia de la velocidad $n^{-1/2}$, la cual tiene lugar para la convergencia de las funciones de distribución. Es un hecho natural, ya que en la estimación del valor de $f(x)$ toma parte, hablando en términos generales, no toda la muestra, sino las observaciones que se han concentrado en cierto entorno decreciente del punto x.

La expresión (12) permite también elegir del modo óptimo la función $q(z)$, o sea, la función para la cual se minimiza Dd^2. Suponiendo, sin limitar la generalidad, que $D = 1$, obtenemos el problema de minimización $d^2 = \int q^2(z)dz$ a condición de que $\int q(z)dz = \int z^2 q(z)dz = 1$, $\int zq(z)dz = 0$.

Nótese que si f tiene derivadas continuas de orden más alto que $2m > 2$, también pueden obtenerse velocidades más altas de convergencia de la diferencia $f_n^*(x) - f(x)$ hacia cero. Sin embargo, en este caso es necesario utilizar las distribuciones generalizadas $\mathbf{Q}$ cuya "densidad" q puede tomar los valores de ambos signos y permite satisfacer las condiciones $\int z^{2m} q(z)dz = 1$, $\int z^j q(z)dz = 0$ para todos los $1 \leqslant j \leqslant 2m - 1$. En este caso, mediante los razonamientos anteriores podemos obtener la velocidad de convergencia de orden de $n^{-\frac{2m}{4m+1}} = n^{-1/2 + \frac{1}{2(4m+1)}}$ la cual será tanto mejor cuanto mayor sea m. Este hecho se explica por la circunstancia de que para $f(x)$ más suaves, en la estimación del valor de $f(x)$ se incorporan los elementos de la muestra, situados en entornos cada vez más amplios del punto x.

Por otro lado, eligiendo funciones suaves $q(z)$, podemos asegurar la posibilidad de estimar no sólo las densidades $f(x)$, sino también sus derivadas. De esto también podemos convencernos a base de los razonamientos anteriormente citados.

La función $f_n^*(x)$, que tiene la forma (4), se llama frecuentemente estimación de Rosenblatt — Parzen de la densidad $f(x)$ o *estimación nuclear de* $f(x)$. En este caso las funciones $q(z)$ se llaman *núcleos*. En la práctica se utilizan a menudo los núcleos "rectangulares", o sea, se supone que

$$q(z) = \begin{cases} 1 \text{ para } z \in [-1/2, \ 1/2], \\ 0 \text{ para } z \notin [-1/2, \ 1/2]. \end{cases}$$

A veces se procede de un modo todavía más sencillo: la recta real se divide en pequeños intervalos Δ_j (de h_n de largo) y se supone que $f_n^*(x) = \frac{\nu_j}{nh_n}$ para $x \in \Delta_j$, donde ν_j es el número de elementos de la muestra que coincidieron con Δ_j. Tal función $f_n^*(x)$ se llama *histograma* de la muestra. Es fácil comprobar que si $f(x)$ es continua, entonces el histograma $f_n^*(x)$, a la par con la función definida en (4), también posee la propiedad de convergencia $f_n^*(x) \xrightarrow[p]{} f(x)$ si $h_n \to 0$, $nh_n \to \infty$.

CAPÍTULO 2

Teoría de estimación de los parámetros desconocidos

El § 2 contiene la descripción de las familias paramétricas más difundidas de distribuciones y sus propiedades principales.

En los §§ 3—6 se exponen métodos principales de obtención de las estimaciones puntuales.

En los §§ 7 y 8 se examinan los enfoques de la comparación de las estimaciones.

Los §§ 9—20 están dedicados a los métodos de construcción de las estimaciones óptimas (en uno u otro sentido). Se destacan las cuatro direcciones siguientes:

1) (§§ 9—11 y 20) Enfoques bayesiano y minimax de la construcción de las estimaciones. Los §§ 9 y 10 son de carácter adicional y contienen las definiciones y la exposición de las propiedades principales de las esperanzas matemáticas condicionales y de las distribuciones condicionales.

2) (§§ 12—15) Construcción de las estimaciones óptimas (eficientes) con ayuda de los principios de suficiencia y de no desplazamiento.

3) (§§ 16, 17 y 22) Construcción de las estimaciones óptimas (eficientes) basándose en la desigualdad de Rao — Cramer.

4) (§§ 18 y 19) Utilización de las consideraciones de invariación.

En los §§ 21—29 se estudian las propiedades asintóticas de la relación de verosimilitud. Sobre esta base se determina la optimación asintótica de las estimaciones de verosimilitud. Los resultados de los §§ 21—29 también constituyen la base de la teoría de los criterios óptimos, desarrollada en el capítulo 3.

Los §§ 31 y 32 están dedicados a la estimación por intervalos.

§ 1. Observaciones preliminares

Como ya hemos señalado en los párrafos precedentes, el objeto inicial de las investigaciones estadísticas está constituido por la muestra

$$X_n = (x_1, \ldots, x_n),\ x_i \in \mathscr{X},$$

de la distribución **P**, la cual es desconocida por completo o parcialmente. En la estadística matemática se destacan, en calidad de principales, las dos siguientes clases de problemas:

1. *Estimación de los parámetros desconocidos.*

2. *Verificación de las hipótesis estadísticas.*

Los problemas de primera clase aparecen cuando por la muestra $X = X_n$ es necesario estimar cualquier característica numérica desconocida θ de la distribución $\mathbf{P}$ (que ya es desconocida). O sea, para la funcional dada

$$\theta = \theta(\mathbf{P}),$$

de la distribución $\mathbf{P}$ debemos señalar la función de la muestra (o bien, que es lo mismo, la estadística)

$$\theta^* = \theta_n^*(X_n)$$

destinada a la utilización, en vez del parámetro θ, en calidad de su aproximación. En el capítulo precedente hemos visto que las premisas para esto existen. La estadística θ^* se llama *estimación* del parámetro θ. Claro está que las estimaciones para el parámetro θ pueden ser muchísimas. El teorema 1.3.1 muestra que, por ejemplo, para la estimación de la funcional $\theta = \theta(\mathbf{P})$, que tiene la forma

$$\theta = \int g(x)\, dF(x),$$

es natural utilizar la estadística

$$\theta^* = \frac{1}{n} \sum_{i=1}^{n} g(x_i).$$

Pero claro que también se pueden examinar otras estimaciones, digamos,

$$\theta^* = \frac{1}{n - \nu_1 - \nu_2} \sum_{j=\nu_1+1}^{n-\nu_2} g(x_{(j)}),$$

donde $x_{(j)}$, $j = 1, \ldots, n$, son los elementos de la serie variacional, etc. En calidad de θ^* también pueden tomarse los valores que no dependen de la muestra. Se puede poner, por ejemplo, $\theta^* \equiv 0$, aunque esto no siempre es racional y es completamente irracional cuando el conjunto de valores posibles de θ no contiene 0.

En relación con la última observación es preciso señalar que en el planteamiento del problema sobre la estimación se indica con frecuencia cuál es el conjunto Θ de los valores posibles de θ. Por ejemplo, si se aprecia la porción θ de un mineral cualquiera contenido en la mena, entonces, claro está que $\theta \in [0, 1]$.

En muchos casos también se sabe de antemano que la distribución $\mathbf{P}$ de la muestra X no puede ser arbitraria, sino que pertenece a una familia determinada de distribuciones $\mathscr{P}$.

Entre los problemas de la estimación de los parámetros figura el ejemplo 1 dado en la Introducción.

Los problemas de segunda clase se refieren a la comprobación de una y otra suposición (hipótesis) sobre la distribución desconocida **P**. Por ejemplo, podemos verificar la hipótesis consistente en que **P** tiene una u otra forma dada. A este tipo de problemas pertenece el ejemplo 2 dado en la Introducción.

Más tarde veremos que no hay diferencia cualitativa entre los problemas de primera clase (teoría de las estimaciones) y de segunda clase (verificación de las hipótesis estadísticas).

En este capítulo expondremos los planteamientos de los problemas y los enfoques que están íntimamente vinculados con los resultados del capítulo precedente y que pueden llamarse "puramente estadísticos" a distinción de los enfoques más generales de la teoría de los juegos, que se examinan en el cap. 5.

Los enfoques puramente estadísticos expresan, en cierta medida, la esencia de los métodos de la estadística matemática. Históricamente tales enfoques fueron comprendidos mucho antes que los métodos más generales. En cuando a su aplicación, por lo visto, el hombre los utilizaba explícita o implícitamente a lo largo de todo el proceso del conocimiento.

Todo esto justifica la exposición independiente de los enfoques puramente estadísticos, a pesar de que ciertos momentos de esta exposición pueden considerarse como casos particulares en el marco de las concepciones más generales. Al mismo tiempo revelaremos cierta insuficiencia del enfoque puramente estadístico para planteamientos más exactos de los problemas. Esto nos ayudará a comprender el carácter racional de otros puntos de vista.

§ 2. Algunas familias paramétricas de distribuciones y sus propiedades

Examinemos algunas familias de distribuciones que dependen de los parámetros (o familias paramétricas de distribuciones) que con frecuencia surgen en los suplementos y que aparecerán en la exposición ulterior tanto de hecho como en calidad de ilustraciones.

1. Distribución normal en una recta. Con el símbolo Φ_{α,σ^2} designaremos la distribución normal con los parámetros (α, σ^2), o sea, la distribución de densidad

$$\varphi_{\alpha,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\alpha)^2}{2\sigma^2}},$$

así que

$$\Phi_{\alpha,\sigma^2}(B) = \int_B \varphi_{\alpha,\sigma^2}(x)dx.$$

Si $\xi \in \Phi_{0,1}$ y $k \geqslant 0$ es un número entero, entonces, evidentemente,

$$\mathbf{M}\xi^{2k+1} = 0.$$

Para los momentos de orden par, utilizando la sustitución $x = \sqrt{2}u$, encontramos

$$\mathbf{M}\xi^{2k} = \frac{2}{\sqrt{2\pi}} \int_0^\infty x^{2k} e^{-x^2/2} dx = \frac{2^{k+1}}{\sqrt{2\pi}} \int_0^\infty u^k e^{-u} \frac{du}{\sqrt{2u}} =$$

$$= \frac{2^k}{\sqrt{\pi}} \Gamma(k + 1/2), \tag{1}$$

donde $\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x} dx$ es la función Γ, $\Gamma(\lambda) = (\lambda - 1)\Gamma(\lambda - 1)$, $\Gamma(1/2) = \sqrt{\pi}$, así que

$$\mathbf{M}\xi^{2k} = (2k - 1)!! = (2k - 1)(2k - 3) \ldots 1.$$

También obtendríamos este resultado si hubiéramos derivado $2k$ veces la función característica $e^{-t^2/2}$ en el punto $t = 0$.

2. Distribución normal multidimensional. En el caso multidimensional $\mathscr{X} = R^m$, el símbolo Φ_{α,σ^2} significará la distribución normal en R^m con el vector de esperanzas matemáticas $\alpha = (\alpha_1, \ldots, \alpha_m)$ y con la matriz de segundos momentos centrales $\sigma^2 = \|\sigma_{ij}\|$, $i, j = 1, \ldots, m$. Si A es la matriz inversa a σ^2 (en los casos cuando ella existe), entonces la densidad $\varphi_{\alpha,\sigma^2}(x)$ en R^m de la distribución Φ_{α,σ^2} tiene la forma a (véase [11], p. 148)

$$\varphi_{\alpha,\sigma^2}(x) = \frac{\sqrt{|A|}}{(2\pi)^{m/2}} \exp\left(-\frac{1}{2}(x - \alpha)A(x - \alpha)^T\right),$$

donde x^T es el vector transpuesto. Recordemos también (ya hemos utilizado este hecho en el § 1.7) que la función característica de la magnitud $\xi \in \Phi_{\alpha,\sigma^2}$ es igual a

$$\mathbf{M}e^{it\xi^T} = \exp\left(it\alpha^T - \frac{1}{2} t\sigma^2 t^T\right),$$

donde $t = (t_1, \ldots, t_m)$ es el vector en R^m.

3. Distribución gamma. El símbolo $\Gamma_{\alpha,\lambda}$ designará la llamada "distribución gamma" (o distribución Γ) con los parámetros (α, λ). La densidad $\gamma_{\alpha,\lambda}(x)$ de esta distribución depende de dos parámetros $\alpha > 0$ y $\lambda > 0$ y es igual a (véase [11] y § 7 del cap. 6)

$$\gamma_{\alpha,\lambda}(x) = \begin{cases} \dfrac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, & x \geqslant 0, \\ 0 & x < 0, \end{cases} \tag{2}$$

donde $\Gamma(\lambda)$ es la función Γ definida en (1). La función característica de la distribución Γ tiene la forma ([11])

$$\int_0^\infty e^{itx}\gamma_{\alpha,\lambda}(x)dx = \left(1 - \frac{it}{\alpha}\right)^{-\lambda}. \tag{3}$$

Si $\xi \in \Gamma_{\alpha,\lambda}$, entonces

$$\mathbf{M}\xi^t = \frac{\alpha^\lambda}{\Gamma(\lambda)}\int_0^\infty x^{\lambda+t-1}e^{-\alpha x}dx = \frac{\alpha^{-t}}{\Gamma(\lambda)}\int_0^\infty y^{\lambda+t-1}e^{-y}dy = \frac{\alpha^{-t}\Gamma(\lambda+t)}{\Gamma(\lambda)}. \tag{4}$$

Para enteros $t > 0$, el mismo resultado podría ser obtenido derivando la función característica. Poniendo $t = 1, 2$, encontramos

$$\mathbf{M}\xi = \lambda/\alpha, \quad \mathbf{D}\xi = \lambda/\alpha^2. \tag{5}$$

De las fórmulas (3) y (4) se deduce que el parámetro α desempeña el papel de escala, así que

$$\eta/\alpha \in \Gamma_{\alpha,\lambda} \quad \text{si} \quad \eta \in \Gamma_{1,\lambda}.$$

En virtud de esta circunstancia, muchas propiedades de la distribución Γ pueden ser estudiadas para un valor cualquiera de α, por ejemplo, para $\alpha = 1$ o para $\alpha = 1/2$. A menudo el segundo valor será para nosotros más cómodo, ya que la distribución $\Gamma_{1/2,\lambda}$ desempeña un importante papel independiente en la estadística matemática y se llama distribución "ji-cuadrado" (o distribución χ^2).

4. Distribución "ji-cuadrado" $\mathbf{H}_k$ con k grados de libertad. Así se denomina la distribución $\mathbf{H}_k = \Gamma_{1/2,\,k/2}$ cuando $k > 0$ son enteros. Conservaremos esta denominación para la distribución $\mathbf{H}_k$ cuando también se trate de $k > 0$ arbitrarios. En virtud de (3), la función característica de la distribución $\mathbf{H}_k$ es igual a

$$(1 - 2it)^{-k/2}.$$

Indiquemos las tres siguientes propiedades de la distribución $\mathbf{H}_k$.

1) *Si η_i son independientes, $\eta_i \in \mathbf{H}_{k_i}$, $i = 1, \ldots, s$, entonces*

$$\sum_{i=1}^{s}\eta_i \in \mathbf{H}_k, \ k = \sum_{i=1}^{s} k_i.$$

Esta propiedad se deduce directamente de la forma de la función característica de la distribución $\mathbf{H}_k$.

2) *Si $\xi \in \Phi_{\alpha,\sigma^2}$, donde Φ_{α,σ^2} es la distribución normal k-dimensional con la matriz no degenerada de segundos momentos σ^2, entonces*

$$Q(\xi) = (\xi - \alpha)\sigma^{-2}(\xi - \alpha)^T \in \mathbf{H}_k.$$

En efecto, la función característica de la variable aleatoria $Q(\xi)$ es igual a

$$\mathbf{M}e^{itQ(\xi)} = \frac{\sqrt{|\sigma^{-2}|}}{(2\pi)^{k/2}} \int \exp\left(-\frac{1}{2}\, Q(x)(1 - 2it)\right) dx_1, \ldots dx_k.$$

Sustituyendo las variables $x_j\sqrt{1 - 2it} = y_j$, obtenemos la expresión

$$(1 - 2it)^{-k/2}\, \frac{\sqrt{|\sigma^{-2}|}}{(2\pi)^{k/2}} \int e^{-\frac{1}{2} Q(y)} dy_1, \ldots dy_k = (1 - 2it)^{-k/2},$$

que es lo que se necesitaba demostrar. El hecho de que la integral en el primer miembro no depende de la variación del dominio de integración se deriva de la analiticidad de la función subintegral y de su decrecimiento rápido cuando $|y| \to \infty$ (compárese con [11], p. 131).

De lo dicho resulta que la distribución $\mathbf{H}_k$ está contenida en la variable aleatoria

$$\chi^2 = \xi_1^2 + \ldots + \xi_k^2,$$

donde ξ_j son independientes, $\xi_j \in \Phi_{0,1}$. El término "número de grados de libertad" está precisamente relacionado con esta representación.

3) Como $\mathbf{M}\xi_1^2 = 1$, $\mathbf{M}\xi_1^4 = 3$, $D\xi_1^2 = 2$ para $\xi_1 \in \Phi_{0,1}$, *entonces, en virtud del teorema central del límite, para* $k \to \infty$,

$$\frac{\chi^2 - k}{\sqrt{2k}} \in \Phi_{0,1}. \tag{6}$$

De aquí y de los teoremas de continuidad enunciados en el § 1.5 se deduce que a la par con (6),

$$\sqrt{2\chi^2} - \sqrt{2k - 1} \in \Phi_{0,1}.$$

Esta convergencia sirve de base para la igualdad aproximada (en caso de k y x grandes) $\mathbf{H}_k(0, x) \approx \Phi(\sqrt{2x} - \sqrt{2k - 1})$, $\Phi(x) = \Phi_{0,1}((-\infty, x))$, la cual, por regla general, resulta más exacta que la aproximación $\mathbf{H}_k((0, x) \approx \Phi\left(\frac{x - k}{\sqrt{2k}}\right)$ que se deduce de (6).

Señalemos otro caso particular de la distribución Γ, el cual aparece a menudo en las aplicaciones.

5. Distribución exponencial. Es la distribución $\Gamma_{\alpha,1}$ de densidad

$$\alpha e^{-\alpha x},\ x > 0.$$

De las fórmulas (5) obtenemos, para $\xi \in \Gamma_{\alpha,1}$,

$$\mathbf{M}\xi = 1/\alpha, \quad D\xi = 1/\alpha^2.$$

Examinemos ahora ciertas distribuciones relacionadas con las distribuciones normal y gamma y que desempeñan un papel importante en la estadística matemática. A distinción de las anteriores, con estas distribuciones no hemos tropezado anteriormente.

6. Distribución de Fisher $\mathbf{F}_{k_1 k_2}$ con k_1 y k_2 número de grados de libertad. Así se llama la distribución de la variable aleatoria

$$\zeta = \eta_1/\eta_2,$$

donde η_j son independientes, $\eta_j \in \mathbf{H}_{k_j}$, $j = 1, 2$. De las propiedades de la distribución Γ se deduce que la distribución de ζ queda igual cuando $\eta_j \in \Gamma_{\alpha,\, k_j/2}$ y para cualquier $\alpha > 0$, y que ζ cuando k_j son enteros, admite la representación

$$\zeta = \frac{\xi_1^2 + \ldots + \xi_{k_1}^2}{\zeta_1^2 + \ldots + \zeta_{k_2}^2},$$

donde las variables aleatorias ξ_j, ζ_k son independientes, $\xi_j \in \Phi_{0,1}$, $\zeta_k \in \Phi_{0,1}$.

Hallemos la densidad de la distribución $\mathbf{F}_{k_1, k_2}$. Tenemos

$$\mathbf{P}(\zeta < x) = \iint\limits_{u/v < x} \Gamma_{1,\lambda_1}(du)\Gamma_{1,\lambda_2}(dv) = \int\limits_{v=0}^{\infty}\int\limits_{u=0}^{vx} \frac{u^{\lambda_1 - 1} v^{\lambda_2 - 1}}{\Gamma(\lambda_1)\Gamma(\lambda_2)} e^{-u-v} du\, dv;$$

$$f_{(\zeta)}(x) = \frac{d\mathbf{P}(\zeta < x)}{dx} = \int\limits_0^{\infty} \frac{(vx)^{\lambda_1 - 1} v^{\lambda_2 - 1}}{\Gamma(\lambda_1)\Gamma(\lambda_2)} e^{-v-vx} v\, dv =$$

$$= \frac{x^{\lambda_1 - 1}}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \int\limits_0^{\infty} v^{\lambda_1 + \lambda_2 - 1} e^{-v(1+x)} dv = \frac{x^{\lambda_1 - 1}\Gamma(\lambda_1 + \lambda_2)}{(1 + x)^{\lambda_1 + \lambda_2}\Gamma(\lambda_1)\Gamma(\lambda_2)}. \qquad (7)$$

Es evidente que la densidad necesaria se obtiene si aquí se sustituye $\lambda_j = k_j/2$. Es fácil determinar los momentos de la variable aleatoria ζ (si éstos existen):

$$\mathbf{M}\zeta^l = \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \int\limits_0^{\infty} \frac{x^{\lambda_1 + l - 1}}{(1 + x)^{\lambda_1 + \lambda_2}} dx = \frac{\Gamma(\lambda_1 + l)\Gamma(\lambda_2 - l)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}. \qquad (8)$$

En particular, cuando $l = 1, 2$, obtenemos

$$\mathbf{M}\zeta = \frac{\lambda_1}{\lambda_2 - 1}, \quad \mathbf{M}\zeta^2 = \frac{\lambda_1(\lambda_1 + 1)}{(\lambda_2 - 1)(\lambda_2 - 2)}.$$

La distribución de Fisher también a veces se llama distribución de Snedecor. Esto se debe al hecho de que Fisher propuso utilizar y tabuló, en

realidad, no la distribución de ζ, sino la de la variable aleatoria $\frac{1}{2}\ln\zeta$. En cuanto a la distribución de ζ, ésta fue tabulada un poco más tarde por Snedecor

7. **Distribución de Student** [*] **T_k con k grados de libertad.** Esta es, por definición, la distribución de la variable aleatoria

$$t = \frac{\xi_0}{\sqrt{\frac{1}{k}(\xi_1^2 + \ldots + \xi_k^2)}},$$

donde ξ_j son independientes, $\xi_j \in \Phi_{0,1}$, $j = 0, \ldots, k$. Es evidente que $-t$ tiene la misma distribución y, por lo tanto, la distribución de Student es simétrica con respecto al origen de coordenadas. Luego

$$t^2 = \frac{k\xi_0^2}{\xi_1^2 + \ldots + \xi_k^2} = \frac{k\eta_1}{\eta_2},$$

donde η_j son independientes, $\eta_1 \in \mathbf{H}_1$, $\eta_2 \in \mathbf{H}_k$. Esto quiere decir que t^2/k tiene la distribución de Fisher. Examinemos la variable aleatoria $\tau = \sqrt{\zeta}$, $\zeta = \eta_1/\eta_2$, $\eta_j \in \mathbf{H}_{k_j}$. Como $\mathbf{P}(\tau < x) = \mathbf{P}(\zeta < x^2)$, la densidad $f_{(\tau)}(x)$ de la variable aleatoria τ será igual a

$$f_{(\tau)}(x) = 2xf_{(\zeta)}(x^2) = 2x\,\frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \cdot \frac{x^{2\lambda_1 - 2}}{(1 + x^2)^{\lambda_1 + \lambda_2}} =$$

$$= \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \cdot \frac{2x^{2\lambda_1 - 1}}{(1 + x^2)^{\lambda_1 + \lambda_2}}, \quad \lambda_j = k_j/2,\ x > 0.$$

De aquí, cuando $\lambda_1 = 1/2$, $\lambda_2 = k/2$, se puede obtener, de un modo evidente, la densidad $|t|/\sqrt{k}$. Como la distribución de t es simétrica, para la densidad $f_{(t)}(x)$ de la variable aleatoria t tenemos finalmente

$$f_{(t)}(x) = \frac{\Gamma((k+1)/2)}{\sqrt{\pi k}\ \Gamma(k/2)}\left(1 + \frac{x^2}{k}\right)^{-(k+1)/2}. \tag{9}$$

Por supuesto que todos los momentos de t de orden impar (si existen) son iguales a cero. Para los momentos de orden par $2l$ tenemos, en virtud de (8),

$$\mathbf{M}t^{2l} = k^l\mathbf{M}\zeta^l = k^l\,\frac{\Gamma(\lambda_1 + l)\Gamma(\lambda_2 - l)}{\Gamma(\lambda_1)\Gamma(\lambda_2)},$$

donde es necesario poner $\lambda_1 = 1/2$, $\lambda_2 = k/2$, $2l < k$. Si $l = 1$ obtenemos

$$\mathbf{M}t^2 = \frac{k}{k-2}.$$

---

[*] *Student* es el seudónimo de W. S. Gosset.

Según su forma, la función $f_{(t)}(x)$ se parece a la densidad de la ley normal. Además, con el crecimiento de k,

$$f_{(t)}(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

que significa la convergencia $t \Leftarrow \Phi_{0,1}$ cuando $k \to \infty$. Sin embargo, $f_{(t)}(x)$ tiene "colas más gruesas", puesto que con el aumento de $|x|$, la función (9) disminuye mucho más lentamente que $e^{-x^2/2}$, así que para todos $b > 0$,

$$\mathbf{T}_k((-b, b)) < \Phi_{0,1}((-b, b)). \tag{10}$$

En este caso, la diferencia entre el segundo y el primer miembro en (10) puede ser considerable cuando k no son grandes.

El lector también puede demostrar la convergencia $t = \sqrt{k}\xi_0/\sqrt{\eta_2}$ hacia la ley normal, utilizando otra vía, por medio del teorema de continuidad. Por ejemplo, basta con notar que $\frac{\eta_2}{k} = \frac{1}{k}(\xi_1^2 + \ldots + \xi_k^2) \underset{c.s.}{\longrightarrow} 1$ y, por lo tanto, $t \underset{c.s.}{\longrightarrow} \xi_0$, $t \Rightarrow \xi_0$.

8. Distribución beta (B-distribución). Así se llama la distribución $\mathbf{B}_{\lambda_1, \lambda_2}$ de densidad

$$f_{(\beta)}(x) = \begin{cases} \dfrac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} x^{\lambda_1 - 1}(1 - x)^{\lambda_2 - 1}, & x \in [0, 1], \\ 0, & x \notin [0, 1] \end{cases}$$

Se denomina así debido a la función beta

$$\mathrm{B}(\lambda_1, \lambda_2) = \int_0^1 x^{\lambda_1 - 1}(1 - x)^{\lambda_2 - 1} dx = \frac{\Gamma(\lambda_1)\Gamma(\lambda_2)}{\Gamma(\lambda_1 + \lambda_2)}.$$

La distribución beta está relacionada con la distribución gamma y la distribución de Fisher por medio de la afirmación siguiente:

Si η_j son independientes, $\eta_j \in \Gamma_{\alpha, \lambda_j}$ (o bien $\eta_j \in \mathbf{H}_{2\lambda_j}$), entonces

$$\beta = \frac{\eta_1}{\eta_1 + \eta_2} = \frac{\zeta}{\zeta + 1} \in \mathbf{B}_{\lambda_1, \lambda_2},$$

donde $\zeta = \eta_1/\eta_2 \in \mathbf{F}_{2\lambda_1, 2\lambda_2}$.

La demostración de esta afirmación es muy fácil, ya que en virtud de (7), $\mathbf{P}(\beta < x) = \mathbf{P}\left(\zeta < \frac{x}{1 - x}\right)$,

$$f_{(\beta)}(x) = f_{(\zeta)}\left(\frac{x}{1-x}\right)\left(\frac{x}{1-x}\right)' = \frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}\left(\frac{x}{1-x}\right)^{\lambda_1-1}\times$$

$$\times(1-x)^{\lambda_1+\lambda_2-2} = \frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}x^{\lambda_1-1}(1-x)^{\lambda_2-1},\ x\in[0,\ 1].$$

Para los momentos de la variable aleatoria β tenemos

$$\mathbf{M}\beta^l = \frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}\int_0^1 x^{\lambda_1+l-1}(1-x)^{\lambda_2-1}dx = \frac{\Gamma(\lambda_1+\lambda_2)\Gamma(\lambda_1+l)}{\Gamma(\lambda_1)\Gamma(\lambda_1+\lambda_2+l)}.$$

Para $l = 1,\ 2$ obtenemos

$$\mathbf{M}\beta = \frac{\lambda_1}{\lambda_1+\lambda_2},\quad \mathbf{M}\beta^2 = \frac{\lambda_1(\lambda_1+1)}{(\lambda_1+\lambda_2)(\lambda_1+\lambda_2+1)}.$$

9. Distribución uniforme. La distribución uniforme sobre $[0,\ 1]$, que se obtiene si se pone $\lambda_1 = \lambda_2 = 1$, es un caso particular de la B-distribución.

Designaremos con el símbolo $\mathbf{U}_{a,b}$ la distribución uniforme sobre el segmento $[a,\ b]$, así que $\mathbf{B}_{1,1} = \mathbf{U}_{0,1}$.

Con ayuda de B-distribución se puede describir la distribución de los términos de la serie variacional $x_{(k)}$ de la muestra X.

Teorema 1. *Si* $X \in \mathbf{P}$ *es la muestra de la distribución* $\mathbf{P}$ *con la función continua de distribución* F, *entonces*

$$y_{(k)} = F(x_{(k)}) \in \mathbf{B}_{k,n-k+1}.$$

Demostración. Como $y_k = F(x_k) \in \mathbf{U}_{0,1}$, entonces $y_{(k)} = F(x_{(k)})$ puede considerarse como término de la serie variacional de la muestra $Y \in \mathbf{U}_{0,1}$. Determinemos $\mathbf{P}(y_{(k)} \in (u, u+du))$. El suceso $\{y_{(k)} \in (u,\ u+du)$ se puede representar como la unión de los sucesos disjuntos

$$A_j = \{y_j \in (u,\ u+du),\ y_j = y_{(k)}\},$$

que se producen cuando y_j adquiere el valor de $(u,\ u+du)$ (esta probabilidad es igual a du), cuando $k-1$ observaciones, de las $n-1$ restantes, caen en el campo de valores de $(0,\ u)$, y cuando $n-k$ observaciones caen en el campo de valores de $(u,\ 1)$. Por consiguiente,

$$\mathbf{P}(A_j) = C_{n-1}^{k-1}u^{k-1}(1-u)^{n-k}du,$$

$$\mathbf{P}(y_{(k)} \in (u,\ u+du)) = nC_{n-1}^{k-1}u^{k-1}(1-u)^{n-k}du.$$

Esto precisamente significa que la densidad $y_{(k)}$ existe y es igual a

$$\frac{n!}{(k-1)!(n-k)!}u^{k-1}(1-u)^{n-k} = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)}u^{k-1}(1-u)^{n-k}. \triangleleft$$

Basándose en el teorema 1 también es fácil obtener la distribución *límite* de los términos de la serie variacional cuando el volumen de la muestra X crece ilimitadamente. Aquí sólo examinaremos un resultado que se deriva de los teoremas de continuidad.

Teorema 2. *Si* $a = \dfrac{k}{n+1} \to a_0 \in (0, 1)$ *cuando* $n \to \infty$, *entonces*

$$y_{(k)} = a + \frac{\sqrt{a_0(1-a_0)}}{\sqrt{n}}\,\xi_n,\quad \xi_n \Leftarrow \Phi_{0,1}.$$

Demostración. En virtud del teorema 1, $y_{(k)} \Leftarrow \mathbf{B}_{k,\,n-k+1}$ y, por lo tanto, en virtud de las propiedades de la B-distribución, es válida la representación

$$y_{(k)} \underset{d}{=} \frac{\eta_1}{\eta_1+\eta_2},\quad \eta_j \Leftarrow \mathbf{H}_{k_j},\ k_1 = 2k,\ k_2 = 2(n-k+1).$$

Pongamos, para comodidad, $a_1 = a$, $a_2 = 1 - a$, y supongamos que $a = a_0$ ha sido fijado. Entonces, evidentemente, $k_j/(n+1) = 2a_j$, $j = 1, 2$ y, en virtud de la propiedad de la distribución χ^2,

$$\eta_j = k_j + \sqrt{2k_j}\,.\ \xi_n^{(j)} \Rightarrow \xi^{(j)} \Leftarrow \Phi_{0,1};$$

$$y_{(k)} = \frac{a_1 + \sqrt{\dfrac{a_1}{n+1}}\,\xi_n^{(1)}}{a_1 + a_2\sqrt{\dfrac{a_1}{n+1}}\,\xi_n^{(1)} + \sqrt{\dfrac{a_2}{n+1}}\,\xi_n^{(2)}}\,.$$

Nos queda utilizar el teorema de continuidad 1.5.3A para

$$H(t) = \frac{t_1}{t_1+t_2},\quad b_n = \frac{1}{\sqrt{n+1}},\quad \eta_n^{(j)} = \sqrt{a_j}\,\xi_n^{(j)}.$$

Como η_j (y, por lo tanto, también $\xi_n^{(j)}$) son independientes y

$$\frac{\partial H}{\partial t_1} = \frac{t_2}{(t_1+t_2)^2},\quad \frac{\partial H}{\partial t_2} = -\frac{t_1}{(t_1+t_2)^2},$$

obtenemos

$$(y_{(k)} - a_1)\sqrt{n+1} \Rightarrow a_2\sqrt{a_1}\ \xi^{(1)} - a_1\sqrt{a_2}\ \xi^{(2)} \underset{d}{=} \sqrt{a_1 a_2}\ \xi,\ \xi \Leftarrow \Phi_{0,1}.$$

Si a depende de n, entonces conviene utilizar la observación 1.5.1. ◁

Corolario 1. *Si* $a = k/(n+1) \to a_0 \in (0, 1)$ *y la función continua F es continuamente derivable en el punto* $\zeta_0 = F^{-1}(a_0)$ *(cuantila de orden* a_0*), entonces*

$$x_{(k)} = \zeta + \frac{\sqrt{a_0(1-a_0)}\ \xi_n}{f(\zeta_0)\sqrt{n}}, \quad \xi_n \in \Phi_{0,1}, \tag{11}$$

donde $\zeta = F^{-1}(a)$ *es una cuantila de orden* a, $f(x) = F'(x)$.

Esta afirmación se obtiene directamente del teorema de continuidad 1.5.3 (teniendo en cuenta la observación 1.5.1) si se utiliza la representación

$$x_{(k)} = F^{-1}(y_{(k)}) = F^{-1}\left(a + \sqrt{\frac{a_0(1-a_0)}{n}}\ \xi_n\right)$$

y el hecho de que $\dfrac{dF^{-1}(x)}{dx} = \dfrac{1}{f(F^{-1}(x))}$.

Observación 1. La afirmación (11) generaliza, de cierto modo, la afirmación del corolario 1.8.1. La misma también puede ser generalizada en otro sentido. Sea, para $x \to \zeta$,

$$|F(x) - F(\zeta)| \sim c\,|x - \zeta|^{\gamma},\ \gamma > 0.$$

Entonces es fácil ver que, cuando $y \to a$,

$$|F^{-1}(y) - F^{-1}(a)| \sim \left|\frac{y-a}{c}\right|^{1/\gamma}$$

y, por lo tanto,

$$(x_{(k)} - \zeta)n^{\frac{1}{2\gamma}} \Rightarrow (a_0(1-a_0))^{\frac{1}{2\gamma}}\,|\xi_c|^{\frac{1}{\gamma}}\ \text{sign}\xi,\ \xi \in \Phi_{0,1} \tag{12}$$

Cuando $\gamma = 1$, $c = f(\zeta)$, de aquí se deduce (11).

10. Distribución de Cauchy $\mathbf{K}_{\alpha,\sigma}$ con parámetros (α, σ). Así se llama la distribución de densidad

$$k_{\alpha,\sigma}(x) = \frac{\sigma}{\pi\,|\sigma^2 + (x-\alpha)^2|} = \frac{1}{\pi\sigma}\cdot\frac{1}{1 + \left(\frac{x-\alpha}{\sigma}\right)^2}.$$

Al igual que en el caso de la ley normal, aquí los parámetros α y σ son, respectivamente, los parámetros de desplazamiento y de escala. La forma de la distribución $\mathbf{K}_{0,1}$ es muy semejante a la de $\Phi_{0,1}$, sin embargo, $k_{0,1}$, al igual que la densidad de la distribución de Student, tiene "colas mucho más gruesas" (o sea, un decrecimiento más lento cuando $|x| \to \infty$), así que la distribución $\mathbf{K}_{0,1}$ no tiene incluso una esperanza matemática finita. En [11] hemos señalado (véase el cap. 7) que las distribuciones $\mathbf{K}_{\alpha,\sigma}$, al igual que las distribuciones normales, poseen propiedad de estabilidad. La función característica $\varkappa_{0,1}(t)$ de la distribución $\mathbf{K}_{0,1}$ es igual a

$$\varkappa_{0,1}(t) = e^{-|t|},$$

por eso $\varkappa_{\alpha,\sigma}(t) = \exp\{i\alpha t - \sigma|t|\}$,

$$\chi_{\alpha_1,\sigma_1}(t)\chi_{\alpha_2,\sigma_2}(t) = \exp\{i(\alpha_1+\alpha_2)t - (\sigma_1+\sigma_2)|t|\},$$

así que la convolución de $\mathbf{K}_{\alpha_1,\sigma_1}$ y $\mathbf{K}_{\alpha_2,\sigma_2}$ es igual a $\mathbf{K}_{\alpha_1+\alpha_2,\sigma_1+\sigma_2}$. No es difícil ver que $\mathbf{K}_{0,1} = \mathbf{T}_1$.

En las aplicaciones se encuentran con frecuencia las funciones de diferente género de las variables aleatorias normalmente distribuidas. Una de ellas es la función exponencial con la cual está relacionada la llamada distribución lognormal.

11. Distribución lognormal $\mathbf{L}_{\alpha,\sigma^2}$. Diremos que $\eta \in \mathbf{L}_{\alpha,\sigma^2}$ si $\ln\eta \in \Phi_{\alpha,\sigma^2}$. En otros términos, $\eta = e^{\xi}$, donde $\xi \in \Phi_{\alpha,\sigma^2}$. De aquí se deduce que la distribución $\mathbf{L}_{\alpha,\sigma^2}$ está concentrada en el semieje positivo.

La densidad de $\eta \in \mathbf{L}_{\alpha,\sigma^2}$, en virtud de las fórmulas para la densidad de la función de la variable aleatoria (véase [11], p. 53), es igual a

$$\varphi_{\alpha,\sigma^2}(\ln x)x^{-1}.$$

Además, hallamos

$$\mathbf{M}\eta = \int e^y \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\alpha)^2}{2\sigma^2}} dy = \exp\frac{(a+\sigma^2)^2-\alpha}{2\sigma^2} \times$$
$$\times \int \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\alpha-\sigma^2)^2}{2\sigma^2}\right) dy = e^{\alpha+\sigma^2/2},$$
$$\mathbf{M}\eta^2 = \int e^2 y \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\alpha)^2}{2\sigma^2}} dy = e^{2\alpha+2\sigma^2}.$$

12. Distribución degenerada. El símbolo $\mathbf{I}_a$ (ya hemos utilizado esta designación en el § 1.2) significará la distribución degenerada concentrada en el punto a.

En el caso general, cuando se examina una *familia arbitraria de distribuciones* que dependen del parámetro θ (escalar o vectorial), utilizaremos la designación $\mathbf{P}_\theta$. La propia familia se designará con el símbolo

$$\{\mathbf{P}_\theta\}_{\theta\in\Theta}$$

donde Θ es el conjunto de valores posibles del parámetro θ. Estas mismas designaciones se emplearán para las familias de distribuciones 1—12. Por ejemplo, $\{\Phi_{\alpha,1}\}_{\alpha\in R}$ significará la familia de todas las distribuciones normales con una varianza unitaria.

Las distribuciones 1—11 son absolutamente continuas con respecto a la medida de Lebesgue. Introduzcamos ahora las designaciones para tres distribuciones *discretas* bien conocidas (absolutamente continuas con respecto a la medida de cálculo $\mu(B)$: $\mu(B) = k$ si B contiene k puntos de valores enteros).

13. Distribución de Bernoulli $\mathbf{B}_p^n$. Según la definición, $\xi \in \mathbf{B}_p^n$ (n es un número entero, $p \in [0, 1]$) si

$$\mathbf{P}(\xi = k) = C_n^k p^k (1 - p)^{n-k}, \quad 0 \leqslant k \leqslant n.$$

14. Distribución de Poisson Π_λ. Esta distribución se determina por medio de la igualdad

$$\Pi_\lambda(B) = \sum_{\substack{k \in B \\ k \geqslant 0}} \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0.$$

15. Distribución polinomial. Designaremos esta distribución por $\mathbf{B}_p^n$, donde $n > 0$ es un número entero, $p = (p_1, \ldots, p_r)$, $p_j \geqslant 0$, $\sum_{j=1}^{r} p_j = 1$. Para el vector aleatorio entero $\nu = \nu_1, \ldots, \nu_r)$ escribiremos $\nu \in \mathbf{B}_p^n$ si para $k = (k_1, \ldots, k_r)$, $k_j \geqslant 0$, $\sum_{j=1}^{r} k_j = n$ es válida la igualdad

$$\mathbf{P}(\nu = k) = \frac{n!}{k_1! \ldots k_r!} p_1^{k_1} \ldots p_r^{k_r}.$$

La distribución $\mathbf{B}_p^n$ corresponde a la sucesión de n pruebas independientes, en cada una de las cuales se produce uno de r casos posibles incompatibles $A_1, \ldots, A_r$; entonces la probabilidad de que aparezca el caso A_j en una prueba es igual a p_j. Las coordenadas ν_j del vector ν significan las frecuencias de aparición de los sucesos A_j después de n pruebas (véase, por ejemplo, [11]). Es evidente que para cada $j = 1, \ldots, r$

$$\nu_j \in \mathbf{B}_{p_j}^n.$$

En el experimento ilustrado, el caso de la j-ésima prueba puede ser descrito por el vector de r-coordenada x_j, cuya $r - 1$ coordenadas son iguales a cero, y una coordenada es igual a 1. El número de esta coordenada es el número del suceso que se produjo en la j-ésima prueba. Evidentemente que $\nu = \sum_{j=1}^{n} \mathrm{x}_j$. Con respecto a la muestra $\underset{\sim}{X}$, formada por $\mathrm{x}_1, \ldots, \mathrm{x}_n$, nos será más cómodo escribir

$$X \in \mathbf{B}_p,$$

donde $\mathbf{B}_p = \mathbf{B}_p^1$. El espacio $\mathscr{X}$ para tal muestra es, por lo visto, finito y consta de r puntos. Si $p = (p_1, p_2)$, $p_1 + p_2 = 1$, obtendremos el esquema de Bernoulli, para el cual utilizaremos las mismas designaciones, identificando $\mathbf{B}(p_1, p_2)$ con $\mathbf{B}_{p_1} = \mathbf{B}_{p_1}^1$ (véase el subpárr.13). En el caso general, la distribución $\mathbf{B}_p$ depende, en realidad, solamente del parámetro de dimensión $r - 1$ $(p_1, \ldots, p_{r-1})$, así que en vez del índice p se podría escribir $(p_1, \ldots, p_{r-1})$.

Muchas de las distribuciones examinadas más arriba, por ejemplo las

distribuciones $\Phi_{0,1}$, H_k, F_{k_1, k_2}, T_k, Π_λ, están tabuladas en los manuales de estadística matemática y se ofrecen en tablas especiales (véase, por ejemplo, [8]).

§ 3. Estimación puntual. Método principal de obtención de estimaciones. Conciliabilidad. Normalidad asintótica

1. Método de sustitución. Conciliabilidad. En el § 1 hemos introducido el concepto de estimación. Formalmente, estimación es lo mismo que estadística, o sea, toda función medible θ^* de una muestra. No formalmente, el sentido que se le da a este término consiste en que llamamos estimaciones θ^* sólo a las estadísticas que deben utilizarse en vez del parámetro desconocido θ. Con otras palabras, θ^* es cierta aproximación para θ, basada en la muestra. La magnitud θ^* también se denomina estimación *puntual* para θ, a distinción de las estimaciones *por intervalo* que serán examinadas más adelante.

La representación de una estimación presupone, de ordinario, la representación de funciones (de la muestra X_n) definidas para todos los valores posibles de n. Por eso, en lo sucesivo el término "estimación" significará la familia de estadísticas $\theta^* = \theta_n^*(X_n)$ definidas para todas los $n = 1, 2, \ldots$, donde θ^* es la función sobre $\mathscr{X}^n$, o bien, que es lo mismo, una función $\theta^* = \theta^*(n, X_\infty)$ definida en el producto del conjunto de números enteros y $\mathscr{X}^\infty$.

De acuerdo con el § 1, consideraremos que en el planteamiento del problema de estimación está definido el conjunto Θ de los posibles valores del parámetro θ y la familia $\mathscr{P}$ de las posibles distribuciones $\mathbf{P}$ de la muestra X (que pueden ser, digamos, sólo las distribuciones normales $\Phi_{\alpha, 1}$ o las distribuciones de Poisson Π_λ para las cuales es preciso estimar los parámetros desconocidos α, λ). Si faltan cualesquiera limitaciones para θ (o para $\mathbf{P}$), entonces podemos considerar que $\theta\mathscr{P}$ coincide con el espacio euclidiano de dimensión correspondiente (con el conjunto de todas las distribuciones).

Si para designar el parámetro, en vez de θ se utiliza otra letra cualquiera, por ejemplo λ, las estimaciones de este parámetro se designarán del mismo modo: añadiendo a λ el índice superior en forma de asterisco. Por ejemplo, para el parámetro α de la ley normal es natural examinar la estimación

$$\alpha^* = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Los momentos muestrales que se utilizan para la estimación

$$\mathbf{M}x_1 = \int x\mathbf{P}(dx) \ \text{y} \ \mathbf{D}x_1 = \int (x - \mathbf{M}x_1)^2\mathbf{P}(dx)$$

tienen sus designaciones especiales tradicionales

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \text{ y } S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Ya hemos señalado que para el parámetro dado se pueden indicar varias estimaciones, tantas como se quiera, y antes de examinar de qué modo en cada situación concreta conviene comparar sus cualidades, fijaremos la atención en ciertos métodos "regulares" generales de su construcción.

Estos métodos agrupan en sí los enfoques más racionales del problema de estimación y posteriormente nos permitirán obtener las mejores estimaciones en uno u otro sentido.

Casi todos los procedimientos de estimación se basan en el siguiente método principal, que podría llamarse *método de sustitución de la distribución empírica* (o simplemente *método de sustitución*).

Sea $X_n \in \mathbf{P}$ y representemos el parámetro desconocido θ en forma de cierta funcional G de la distribución $\mathbf{P}$:

$$\theta = G(\mathbf{P}) \cdot$$

Supongamos, luego, que $\mathbf{P}_n^*$ significa, como antes, la distribución empírica. Entonces, el método de sustitución prescribe que en calidad de estimación θ^* se tome la función

$$\theta^* = G(\mathbf{P}_n^*).$$

Tales estimaciones serán llamadas *estimaciones por el método de sustitución* o simplemente *estimaciones de sustitución.*

La funcional G se da, a veces, en forma implícita como solución de cierta ecuación $H(\theta, \mathbf{P}) = 0$, resoluble con respecto a θ. En este caso, en consonancia con la definición principal, llamaremos estimaciones de sustitución a toda solución de la ecuación $H(\theta, \mathbf{P}_n^*) = 0$.

Si se sabe que el conjunto de los posibles valores del parámetro $\theta \in \in R^k$ está limitado por el dominio Θ de R^k, el cual no coincide con R^k, esta información se puede tener en cuenta al construir las estimaciones de sustitución. Admitamos que el domino Θ está cerrado y sea $\mathcal{P}$ el conjunto de las posibles distribuciones de la muestra X, $\Theta = \{G(\mathbf{P})\}_{\mathbf{P} \in \mathcal{P}}$. Definamos la funcional $G_1(\mathbf{P})$ para $\mathbf{P}$ arbitraria, como el valor de $t \in \Theta$ para el que se alcanza

$$\min_{t \in \Theta} |t - G(\mathbf{P})| = |G_1(\mathbf{P}) - G(\mathbf{P})|, \qquad (1)$$

así que $G_1(\mathbf{P})$ es el punto de Θ más próximo a $G(\mathbf{P})$. Como $G_1(\mathbf{P}) = G(\mathbf{P}) = \theta$, si $\mathbf{P} \in \mathcal{P}$, entonces la estimación

$$\theta^* = G_1(\mathbf{P}_n^*), \qquad (2)$$

junto con $G(\mathbf{P}_n^*)$, será la estimación de sustitución, con la particularidad de que el conjunto de los posibles valores de θ^* pertenecerá a Θ.

En cuanto a las estimaciones (1) y (2) diremos que se han obtenido debido a la *contracción del método de sustitución.*

Supongamos, por ejemplo, que se estima el parámetro α de la distribución normal $\Phi_{\alpha,1}$ y que sabemos de antemano que $\alpha \in [0, 1]$. Entonces puede resultar que la estimación $\alpha^* = \bar{x} \notin [0, 1]$ (evidentemente que $\bar{x} = \int t dF_n^*(t)$ es la estimación de sustitución). La contracción del método de sustitución recomienda en calidad de estimación tomar el punto $[0, 1]$ más próximo a $\bar{x}$.

Señalemos ahora, que en la forma enunciada, el método de sustitución no siempre tiene sentido. El hecho consiste en que la funcional inicial G puede resultar no definida sobre el conjunto de distribuciones empíricas. Supongamos, por ejemplo, que es sabido de antemano que la distribución $\mathbf{P}$ pertenece a la clase $\mathscr{P}$ de distribuciones absolutamente continuas con respecto a la medida de Lebesgue, así que cada $\mathbf{P} \in \mathscr{P}$ tiene una densidad igual a f.

Pero a nosotros nos interesa el valor de

$$\theta = G(\mathbf{P}) = \int f^2(x)dx = \int \left(\frac{d\mathbf{P}}{dx}\right)^2 dx.$$

Está claro que en este caso $G(\mathbf{P}_n^*)$ no tiene sentido, ya que $\mathbf{P}_n^*$ es una distribución discreta. En tales casos el método de sustitución siempre puede ser modificado naturalmente de manera que conserve su esencia. En el ejemplo citado, donde $G(\mathbf{P})$ es la funcional de la densidad f, conviene, en calidad de θ^*, examinar, de acuerdo con el método de sustitución, el valor de $G(\mathbf{P}_n^{**})$, donde $\mathbf{P}_n^{**}$ es la distribución empírica suavizada (véase el § 1.10) que asegura la convergencia de la densidad empírica hacia $f(x)$.

También puede resultar que en algunos casos $G(\mathbf{P}_n^*)$ tenga sentido no para todas las X_n, sino sólo para $X_n \in A_n$, donde $\mathbf{P}(X_n \in A_n) \to 1$ cuando $n \to \infty$. Esta circunstancia no tendrá ninguna importancia en cuanto a la esencia de la exposición ulterior del material, y para precisar podemos poner $G(\mathbf{P}_n^*) = 0$ para $X_n \notin A_n$. En este párrafo, para simplificar, estimaremos que $G(\mathbf{P}_n^*)$ tiene sentido para todas $X_n \in \mathscr{X}^n$, y que θ^* es una variable aleatoria, o sea, que la función $G(\mathbf{P}_n^*)$ realiza la aplicación medible de $\mathscr{X}^n$ en R^k, donde k es la dimensión de θ.

El principio de sustitución es un enfoque muy natural del problema, puesto que, como ya sabemos, la distribución $\mathbf{P}_n^*$ se aproxima ilimitadamente a $\mathbf{P}$ a medida que crece n.

Sea $X_n = [X_\infty]_n$.

Definición 1. La estimación $\theta^* = \theta_n^*(X_n)$ (o la sucesión $\theta_n^*(X)$) se llama

conciliable si

$$\theta^* \underset{P}{\to} \theta$$

cuando $n \to \infty$.

La estimación θ^* se denomina *fuertemente conciliable* si, para $n \to \infty$,

$$\theta^* \underset{c.s}{\to} \theta$$

Sea F, como siempre, la función de distribución correspondiente a $\mathbf{P}$.

Teorema 1. *Supongamos que $\theta = G(\mathbf{P})$ y que la funcional G pertenece a una de las dos clases, o que es representable en la forma*

$$G(\mathbf{P}) = h\left(\int g(x)dF(x)\right), \tag{I}$$

donde h es una función continua en el punto $a = \int g(x)dF_0(x)$ (funcional de tipo I), o representable en la forma

$$G(\mathbf{P}) = G_1(F), \tag{II}$$

donde la funcional G_1 es continua en el punto F_0, en la métrica uniforme (funcional de tipo II). *Entonces, si $X \in F_0$, $\theta^* = G(\mathbf{P}_n^*)$ es una estimación fuertemente conciliable:*

$$\theta^* \underset{c.s.}{\longrightarrow} \theta.$$

La afirmación de este teorema se deduce directamente del teorema 1.4.1.

2. Normalidad asintótica. Caso unidimensional.

Definición 2. La estimación θ^* del parámetro θ se llama *asintóticamente normal* (*a.n.*) *con coeficiente* $\sigma^2 \geqslant 0$, si $(\theta^* - \theta)\sqrt{n} \in \Phi_{0,\sigma^2}$.

La última relación también puede leerse del modo siguiente: la estimación θ^* a.n. con los parámetros $(\theta, \sigma^2/n)$.

Supongamos que θ^* es la estimación de sustitución del parámetro $\theta = G(\mathbf{P})$ y que se cumple (I), o sea, que

$$\theta^* = h\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right) \tag{3}$$

es una estadística de tipo I. Entonces, de los resultados del § 1.7 se deduce la afirmación siguiente. Supongamos que θ es un parámetro escalar, y g, una función escalar.

Teorema 2. *Sea $X \in F_0$, h derivable en el punto $a = \int g(x)dF_0(x)$,* $0 < |h'(a)| < \infty$, $\int g^2(x)dF_0(x) < \infty$. *Entonces θ^* es la estimación a.n. con coeficiente*

$$\sigma^2 = [h'(a)]^2 \int (g(x) - a)^2 dF_0(x).$$

Los ejemplos examinados en el § 1.7 también pueden utilizarse como ilustraciones de este teorema, ya que las estadísticas examinadas en ellos se utilizan en calidad de estimaciones.

Análogamente podríamos, utilizando los resultados del § 1.8, obtener las condiciones de normalidad asintótica de las estimaciones que son estadísticas de tipo II. El lector puede obtener las afirmaciones necesarias, utilizando el teorema 1.8.1 sin cualesquiera modificaciones, pero exigiendo, no obstante, que en su enunciación se cumpla $k = 1$, y que la derivada g sea tal que $g(F_0, w^\circ) \in \Phi_{0,\sigma^2}$.

3. Normalidad asintótica. Caso de parámetro multidimensional.

Definición 2A. La estimación $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ se denomina *estimación a.n $\theta = (\theta_1, \ldots, \theta_k)$ con matriz σ^2, si*

$$(\theta^* - \theta)\sqrt{n} \in \Phi_{0,\sigma^2}, \tag{4}$$

donde Φ_{0,σ^2} es la distribución normal k-dimensional con vector nulo de las esperanzas matemáticas y con matriz de segundos momentos $\sigma^2 = \|\sigma_{ij}\|$. La densidad de esta distribución es igual (véase el § 2) a

$$\Phi_{0,\sigma^2}(x) = \frac{\sqrt{|A|}}{(2\pi)^{k/2}} e^{-\frac{1}{2} xAx^T},$$

donde A es una matriz inversa a σ^2, $x = (x_1, \ldots, x_k)$.

Si θ^* es la estimación de la sustitución y la misma es una estadística de tipo I (o sea, representable en forma de (3), donde g, hablando en general, junto con θ^* y h, es una función vectorial), entonces, para determinar las condiciones de normalidad asintótica se puede utilizar el teorema 1.7.1A y la observación a él. En este caso obtenemos la afirmación siguiente.

Teorema 2A. *Supongamos que $\theta^* \in R^k$ se define por la igualdad* (1), *donde $g = (g_1, \ldots, g_s) \in R^s$, y la función vectorial $h(t) = (h_1(t), \ldots, h_k(t))$, $t = (t_1, \ldots, t_s)$ tiene en el punto $a = (a_1, \ldots, a_s)$, $a_j = \int g_j(x)dF_0(x)$ las derivadas parciales $\frac{\partial h_l}{\partial t_j}(a)$, $l = 1, \ldots, k$, $j = 1, \ldots, s$. Entonces, si $X \in F_0$,*

$$(\theta^* - \theta)\sqrt{n} \Rightarrow \xi H^T,$$

donde $\xi = (\xi_1, \ldots, \xi_s) \in \Phi_{0,d^2}$ es el vector normalmente distribuido, con la media nula y la matriz de segundos momentos $d^2 = \|d_{ij}\|$, $d_{ij} = \mathbf{M}(g_i(x_1) - a_i)(g_j(x_1) - a_j)$, $i, j = 1, \ldots, s$; $H = \|h_{ij}\|$ es una matriz de dimensión $k \times s$, con los elementos $h_{ij} = \frac{\partial h_i}{\partial t_j}(a)$, $i = 1, \ldots, k$; $j = 1, \ldots, s$.

Esto significa, a su vez, que al cumplirse las condiciones del teorema 2A, θ^* es una estimación a.n. con matriz $\sigma^2 = Hd^2H^T = \mathbf{M}H\xi^T\xi H^T$. Cabe

señalar que las matrices σ^2 y d^2 aquí tienen, hablando en general, dimensiones diferentes (k y s).

§ 4. Realización del método de sustitución en el caso paramétrico. Método de momentos

Sea $X \in \mathbf{P}_\theta$, donde $\{\mathbf{P}_\theta\}_{\theta \in \Theta}$ es la familia de distribuciones $\mathbf{P}_\theta$ que ya conocemos y que dependen del parámetro θ. En nuestras investigaciones, el parámetro desconocido θ del conjunto Θ puede ser tanto escalar como vectorial. Por ejemplo, si $X \in \Phi_{\alpha,\sigma^2}$, entonces $\theta = (\alpha, \sigma^2)$ es bidimensional, y el conjunto Θ puede ser tanto un semiplano $\{-\infty < \alpha < \infty, \sigma \geqslant 0\}$ como cualquier parte de éste.

La esperanza matemática y la varianza de la estadística $S = S(X)$ en función de la distribución $\mathbf{P}_\theta$ serán designadas por $\mathbf{M}_\theta S$ y $\mathbf{D}_\theta S$, respectivamente.

Más adelante examinaremos algunos métodos de estimación, cada uno de los cuales puede interpretarse como la realización del principio de sustitución de una distribución empírica.

1. Método de momentos. Caso unidimensional. Escojamos $g(x)$ de tal modo que la función

$$m(\theta) = \mathbf{M}_\theta g(\mathrm{x}_1) = \int g(x)\mathbf{P}_\theta(dx) \tag{1}$$

sea monótona y continua. El campo $m(\Theta)$ de valores $m(\theta)$, $\theta \in \Theta$ tiene la misma "naturaleza" que Θ. Si, por ejemplo, Θ es un segmento del eje real, $m(\Theta)$ también será un segmento.

Es evidente que la ecuación $m(\theta) = t$ es unívoca y continuamente resoluble en el campo $m(\Theta)$ respecto a $\theta{:}\theta = m^{-1}(t)$, y que (1) se puede escribir del modo equivalente en la forma

$$\theta = m^{-1}\left(\int g(x)\mathbf{P}_\theta(dx)\right). \tag{2}$$

Supongamos simplemente, que

$$\bar{g} = \int g(x)d\mathbf{P}_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} g(\mathrm{x}_i) \in m(\Theta)$$

para todas $X \in \mathscr{X}^n$.

Definición 1. Se llama estimación por el *método de momentos* la estimación

$$\theta^* = m^{-1}(\bar{g}).$$

Si $\bar{g} \notin m(\Theta)$, se puede poner, conforme a (3.1) y (3.2),

$$\theta^* = m^{-1}(\bar{g}_0),$$

donde $\bar{g}_0 \in m(\Theta)$ es el punto de $m(\Theta)$ más próximo a $\bar{g}$.

No es difícil darse cuenta que esto constituye la estimación con arreglo al principio de sustitución. La elección de la función $m(\theta)$ nos ha permitido expresar θ en forma de la funcional (2). También está claro que la estimación (3) es una estadística de tipo I, así que, en virtud del teorema 3.1, las estimaciones conforme al método de momentos serán fuertemente conciliables. Si además, la función m es derivable en el punto θ, $\int g^2(x)\mathbf{P}_\theta(dx) < \infty$, entonces, según el teorema 3.2, la estimación con arreglo al método de momentos será a.n. con coeficiente $(m'(\theta))^{-2}\mathbf{D}_\theta g(x_1)$.

El método de momentos fue propuesto por C. Pearson (en forma algo más particular) e históricamente es el primer método regular para construir estimaciones.

La propia denominación de "método de momentos" se debe al hecho de que su esencia consiste en igualar entre sí los momentos "teóricos" y empíricos (esperanzas matemáticas) de la magnitud $g(x_1)$: pues la estimación (3) no es otra cosa sino la solución de la ecuación

$$m(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(x_i). \tag{4}$$

También se puede añadir que en calidad de $g(x)$ se elige con frecuencia la función $g(x) = x$ o bien $g(x) = x^k$, $k > 1$, así que nuestra ecuación se convierte en ecuación para momentos ordinarios.

La igualdad (4) también puede considerarse como el resultado de la igualación del valor medio de la magnitud $g(x_1)$ "en el espacio", a su valor medio "en el tiempo".

El carácter no únivoco del método de momentos, así como de todo el principio de sustitución, aquí se manifiesta sobre todo bien: pues casi nada limita la elección de la función $g(x)$.

Ejemplo 1. Supongamos que $X \in \mathbf{\Gamma}_{\alpha,1}$ y que α se desconoce. Construyamos las estimaciones conforme al método de momentos con dos funciones elementales $g(x)$: $g_1(x) = x$ y $g_2(x) = x^2$. Son válidas las igualdades siguientes (véase el punto 5 del § 2)):

$$m_1(\alpha) = \mathbf{M}_\alpha g_1(x_1) = \int_0^\infty x\Gamma_{\alpha,1}(dx) = 1/\alpha,$$

$$m_2(\alpha) = \mathbf{M}_\alpha g_2(x_1) = \int_0^\infty x^2\Gamma_{\alpha,1}(dx) = 2/\alpha^2.$$

Resolviendo las ecuaciones $m_1(\alpha) = \bar{x}$, $m_2(\alpha) = \frac{1}{n} \sum_{i=1}^{n} x_i^2$, obtenemos

las estimaciones según el método de momentos

$$\alpha^* = (\bar{X})^{-1} \text{ y } \alpha^{**} = \left(\frac{1}{2n}\sum_{i=1}^{n} x_i^2\right)^{-1/2}, \tag{5}$$

Estas dos estimaciones son estadísticas de tipo I y podemos describir sus propiedades asintóticas. A base de las igualdades (2.4) obtenemos

$$D_\alpha g_1(x_1) = \mathbf{D}_\alpha x_1 = 1/\alpha^2, \ \mathbf{D}_\alpha g_2(x_1) = \mathbf{D}_\alpha x_1^2 = 20/\alpha^4.$$

En vista de que para la primera estimación, $m_1'(\alpha) = -1/\alpha^2$, y para la segunda, $m_2'(\alpha) = -4/\alpha^3$, a base de los teoremas 3.1, 3.2 obtenemos que ambas estimaciones α^* y α^{**} son fuertemente conciliables y a.n. con coeficientes, respectivamente,

$$\frac{1}{\alpha^2} \cdot \alpha^4 = \alpha^2, \ \frac{20}{\alpha^4} \cdot \frac{\alpha^2}{16} = \frac{5}{4}\alpha^2.$$

Evidentemente, conviene dar preferencia a α^*, ya que su "dispersión", en caso de grandes valores de n alrededor del valor verdadero de α, que se mide con arreglo a la varianza de la distribución límite, es menor que la "dispersión" para α^{**}.

2. Método de momentos. Caso multidimensional. De un modo completamente análogo se examina el caso cuando θ es un parámetro multidimensional.

Supongamos, como antes, que k es la dimensión de θ. Elijamos la función vectorial $g(x) = (g_1(x), \ldots, g_k(x))$ de modo que la ecuación

$$m(\theta) = t,$$

donde $t = (t_1, \ldots, t_k)$, $m(\theta) = (m_1(\theta), \ldots, m_k(\theta))$,

$$m_j(\theta) = \mathbf{M}_\theta g_j(x_1) = \int g_j(x)\mathbf{P}_\theta(dx),$$

sea unívoca y continuamente resoluble con respecto a $\theta = m^{-1}(t)$ en el campo $m(\Theta)$ de valores $m(\theta)$, $\theta \in \Theta$. Admitamos simplemente, que el vector

$$\bar{g} = \left(\frac{1}{n}\sum_{i=1}^{n} g_1(x_i), \ \ldots, \ \frac{1}{n}\sum_{i=1}^{n} g_k(x_i)\right)$$

pertenece al campo $m(\Theta)$ de todas $X \in \mathscr{X}^n$.

Definición 1A. La estimación $\theta^* = m^{-1}(\bar{g})$ se llama *estimación por el método de momentos.*

Como antes, del teorema 3.1 se deduce que tales estimaciones θ^* serán fuertemente conciliables.

Para que tenga lugar θ^* a.n. es necesario exigir adicionalmente que la función m sea derivable, $\int g_j^2(x)\mathbf{P}_\theta(dx) < \infty$. La afirmación acerca de la distribución límite de θ^* se obtiene fácilmente con ayuda del teorema 3.2A.

Ejemplo 2. Examinemos en calidad de $\{\mathbf{P}_\theta\}$ la familia de distribuciones normales Φ_{α,σ^2}. Suponiendo $g_1(x) = x$, $g_2(x) = x^2$, obtenemos las ecuaciones siguientes para el método de momentos:

$$\alpha = \bar{x}, \ \sigma^2 + \alpha^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2,$$

cuya solución es

$$\alpha^* = \bar{x}, \ (\sigma^2)^* = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - (\bar{x})^2 = S^2.$$

Proponemos al lector, en calidad de ejercicio, hallar, basándose en el método de momentos, las estimaciones para todas las familias paramétricas expuestas en el § 2.

3. Método generalizado de momentos. Es posible la siguiente generalización del método de momentos, la cual amplía considerablemente la clase de estimaciones antes examinada. Limitémonos simplemente al caso del parámetro unidimensional θ. Examinemos la función de dos variables $g(x, \theta)$ y supongamos que para toda distribución $\mathbf{P}$ la ecuación

$$\int g(x, \theta)\mathbf{P}(dx) = \int g(x, \theta)\mathbf{P}_\theta(dx) \tag{6}$$

es resoluble con respecto a $\theta = G(\mathbf{P})$, de modo que la última igualdad, junto con (6), se convierta en la identidad $\theta = G(\mathbf{P}_\theta)$ cuando $\mathbf{P} = \mathbf{P}_\theta$.

Llamaremos *estimación por el método generalizado de momentos*, la estimación

$$\theta^* = G(\mathbf{P}_n^*).$$

Es evidente que, al igual que las estimaciones por el método de momentos, éstas son estimaciones de sustitución. La investigación de las propiedades de tales estimaciones es más difícil. De esto nos convenceremos en los párrafos sucesivos, puesto que resultará que una de las estimaciones de sustitución que estudiaremos detalladamente será la estimación por el método generalizado de momentos.

§ 5*. Método de distancia mínima

El método indicado en el título, al igual que el de momentos, es la realización del principio de sustitución y consiste en lo siguiente. Examinemos cualquier funcional de dos distribuciones $d(\mathbf{P}, \mathbf{Q})$, la cual posee la propie-

dad consistente en que como función de $\mathbf{Q}$ dicha funcional alcanza su valor mínimo cuando $\mathbf{Q} = \mathbf{P}$ y $d(\mathbf{P}, \mathbf{Q}) > d(\mathbf{P}, \mathbf{P})$ cuando $\mathbf{Q} \neq \mathbf{P}$. Vamos a considerar la magnitud $d(\mathbf{P}, \mathbf{Q})$ (o bien $d(\mathbf{P}, \mathbf{Q}) - d(\mathbf{P}, \mathbf{P})$) como la "distancia" entre $\mathbf{Q}$ y $\mathbf{P}$, de modo que $\mathbf{P}$ se pueda determinar como el valor de $\mathbf{Q}$ con el que $d(\mathbf{P}, \mathbf{Q})$ alcanza su valor mínimo.

Supongamos ahora que $X \in \mathbf{P}$, $\mathbf{P}$ se desconoce y pertenece a la familia $\mathscr{P}$. Designemos por $(\mathbf{Q})_{\mathscr{P}}$ la distribución de $\mathscr{P}$ inmediata a la distribución $\mathbf{Q}$ en sentido de la distancia d, y supongamos que ella existe:

$$d((\mathbf{Q})_{\mathscr{P}}, \mathbf{Q}) = \min_{\Pi \in \mathscr{P}} d(\Pi, \mathbf{Q}),$$

así que $(\mathbf{Q})_{\mathscr{P}} = \mathbf{Q}$ si $\mathbf{Q} \in \mathscr{P}$.

Definición 1. Se llama *estimación de la distribución* $\mathbf{P}$ *conforme al valor mínimo de la distancia d*, la distribución $\mathbf{P}^* = (\mathbf{P}_n^*)_{\mathscr{P}} \in \mathscr{P}$, donde $\mathbf{P}_n^*$ es, como antes, la distribución empírica.

Ahora bien, cuando $\Pi = \mathbf{P}^* = (\mathbf{P}_n^*)_{\mathscr{P}}$ se minimiza $d(\Pi, \mathbf{P}_n^*)$. Si $\mathscr{P}$ coincide con el conjunto de todas las distribuciones, es evidente que $\mathbf{P}^* = \mathbf{P}_n^*$.

Supongamos ahora que $\mathscr{P} = \{\mathbf{P}_\theta\}_{\theta \in \Theta}$ es una familia paramétrica que satisface la condición siguiente:

$$A_0 \qquad \mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_2} \text{ cuando } \theta_1 \neq \theta_2.$$

En este caso la aplicación de $\theta \to \mathbf{P}_\theta$ es biunívoca, por eso la distribución $\mathbf{P} \in \mathscr{P}$ permite restablecer únicamente el parámetro θ con el que $\mathbf{P} = \mathbf{P}_\theta$. Este hecho también puede expresarse de otra manera: existe la funcional G definida sobre $\mathscr{P}$ de tal modo que $\theta = G(\mathbf{P}_\theta)$.

Introduzcamos en este planteamiento la funcional $G_1(\mathbf{Q}) = G((\mathbf{Q})_{\mathscr{P}})$ que es, evidentemente, el valor de $\theta \in \Theta$ con el que $\mathbf{P}_\theta$ será la distribución inmediata a $\mathbf{Q}$ en sentido de la distancia d, así que

$$G_1(\mathbf{P}_\theta) = G(\mathbf{P}_\theta) = \theta. \tag{1}$$

Definición 2. La estimación $\theta^* = G_1(\mathbf{P}_n^*)$ se denomina *estimación del parámetro θ por el valor mínimo de la distancia d.*

En otros términos, θ^* es el valor de Θ con el que

$$d(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*) = \inf_{\theta \in \Theta} d(\mathbf{P}_\theta, \mathbf{P}_n^*).$$

Es evidente que aquí otra vez tropezamos con el principio de sustitución. Esto se deduce de las definiciones y de (1). Claro está que la distancia d y la familia $\mathscr{P} = \{\mathbf{P}_\theta\}$ deben poseer propiedades capaces de asegurar la mensurabilidad de la aplicación de $\mathscr{X}^n$ en R^k, que se realiza mediante la funcional $G_1(\mathbf{P}_n^*)$, de modo que θ^* sea una variable aleatoria.

Ahora señalemos que *en el caso paramétrico, al cumplirse la condición*

(A_0), *la contracción del método de sustitución* (véanse (3.1) y (3.2)) *y el método de distancia mínima proporcionan la misma clase de estimaciones.*

En efecto, ya sabemos que las estimaciones de distancia mínimas θ^* son las estimaciones por el método de sustitución, en este caso $\theta^* \in \Theta$. Supongamos ahora que θ^* es la estimación por el método de sustitución $\theta^* = G(\mathbf{P}_n^*)$, donde $G(\mathbf{P}_\theta) = \theta$, $\theta^* \in \Theta$. Determinemos la distancia $d(\mathbf{P}, \mathbf{Q}) = |G(\mathbf{P}) - G(\mathbf{Q})|$. Entonces, evidentemente, para $\theta = \theta^*$ se alcanza

$$\inf_{\theta \in \Theta} d(\mathbf{P}_\theta, \mathbf{P}_n^*) = \inf_{\theta \in \Theta} |G(\mathbf{P}_\theta) - G(\mathbf{P}_n^*)| = \inf_{\theta \in \Theta} |\theta - G(\mathbf{P}_n^*)| = 0.$$

También se puede notar que el método de momentos es mucho más estrecho que el de sustitución, puesto que es evidente que no cada funcional G tal que $G(\mathbf{P}_\theta) = \theta$, admite la representación de la forma

$$G(\mathbf{P}_\theta) = m^{-1}\left(\int g(x)\mathbf{P}_\theta(dx)\right).$$

Volvamos a las estimaciones de distancia mínima. Está claro que se pueden señalar muchas distancias "racionales" d que pueden utilizarse para construir las estimaciones. Podríamos, en calidad de d, tomar la distancia

$$d(\mathbf{P}, \mathbf{Q}) = \sup_x |F_P(x) - F_Q(x)|$$

o bien

$$d(\mathbf{P}, \mathbf{Q}) = \int (F_P(x) - F_Q(x))^2 dF_Q(x),$$

donde $F_P(x)$ es la función de distribución que corresponde a la distribución $\mathbf{P}$. Aquí serán estimaciones θ^* por la distancia mínima los valores de θ con los que se alcanza, respectivamente,

$$\inf_\theta \sup_x |F_{P_\theta}(x) - F_n^*(x)|, \tag{2}$$

$$\inf_\theta \int (F_{P_\theta}(x) - F_n^*(x))^2 dF_n^*(x) = \inf_\theta \frac{1}{n} \sum_{k=1}^{n} \left(F_{P_\theta}(x_{(k)}) - \frac{k-1}{n}\right)^2.$$

En algunos problemas (compárese esto con [48]) se utilizan las llamadas estimaciones conforme al valor mínimo de χ^2 (ji-cuadrado). Son las estimaciones con arreglo al valor mínimo de la distancia

$$d(\mathbf{P}, \mathbf{Q}) = \sum_{i=1}^{r} \frac{(\mathbf{P}(\Delta_i) - \mathbf{Q}(\Delta_i))^2}{\mathbf{P}(\Delta_i)},$$

donde $\Delta_1, \ldots, \Delta_r$ es la partición de R (o bien de R^m si x_j son m-dimensionales) en $r < \infty$ intervalos, así que $\bigcup_{i=1}^{r} \Delta_i = R$. Ahora bien, la

estimación θ^* conforme al valor mínimo de χ^2 es el valor de θ con el que se minimiza

$$n\sum_{i=1}^{r}\frac{(\mathbf{P}_\theta(\Delta_i)-\nu_i/n)^2}{\mathbf{P}_\theta(\Delta_i)}=\sum_{i=1}^{r}\frac{(n\mathbf{P}_\theta(\Delta_i)-\nu_i)^2}{n\mathbf{P}_\theta(\Delta_i)}. \qquad (3)$$

Aquí $\nu_i = n\mathbf{P}_n^*(\Delta_i)$ es el número de observaciones x_j que adquirieron los valores del intervalo Δ_j. La estadística en el segundo miembro (3) es la estadística χ^2 que ya conocemos, de aquí precisamente procede la denominación de dicha estimación.

Más adelante veremos que existe tal funcional G, $\theta = G(\mathbf{P}_\theta)$ con la que las estimaciones según el principio de sustitución, llamadas estimaciones de verosimilitud máxima, serán las mejores en cierto sentido. En virtud de esta circunstancia, las estimaciones examinadas en este párrafo no tienen, hablando en general, mucha aplicación y por eso no merece la pena detenerse más en ellas.

§ 6. Método de verosimilitud máxima

Otra vez supongamos que $\mathcal{P}$ es una familia paramétrica $\{\mathbf{P}_\theta\}_{\theta\in\Theta}$. En lo sucesivo, con arreglo a esta familia admitiremos, por doquier donde sea necesario, que está cumplida la condición

$$(A_0) \qquad \mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_2} \quad \text{cuando} \quad \theta_1 \neq \theta_2,$$

así como la condición siguiente, que llamaremos condición (A_μ).

(A_μ): *en el espacio de fase* $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$ *existe una medida* σ*-finita* μ *tal que todas las distribuciones* $\mathbf{P}_\theta \in \mathcal{P}$ *tienen, respecto a esta medida, la densidad* $f_\theta(x) = \frac{d\mathbf{P}_\theta}{d\mu}(x)$, *así que*

$$\mathbf{P}_\theta(B) = \int_B f_\theta(x)\mu(dx).$$

En este caso se dice que la medida μ *domina* las distribuciones $\mathbf{P}_\theta$.

Todas las familias de distribuciones examinadas en el § 2 satisfacen, evidentemente, las condiciones (A_0) y (A_μ). Para ciertas distribuciones, en calidad de μ es necesario adoptar la medida de Lebesgue (distribuciones absolutamente continuas), y para otras, la medida de cálculo (distribuciones discretas). La medida de cálculo μ se define así: $\mu(B) = k$, donde k es el número de puntos con coordenadas de valores enteros pertenecientes a B.

A las primeras pertenecen las distribuciones normal Φ_{α,σ^2}, lognormal $\mathbf{L}_{\alpha,\sigma^2}$, las distribuciones Γ y $\mathbf{B}$, la distribución uniforme, la distribución

de Cauchy y las distribuciones de Student y de Fisher, y a las segundas, las distribuciones de Bernoulli y Poisson, así como las distribuciones degeneradas en cero y polinomiales. La forma de densidades $f_\theta(x)$ de estas distribuciones se da en el § 2. En el caso discreto (cuando μ es la medida de cálculo), la densidad $f_\theta(x)$ coincide con la probabilidad $\mathbf{P}_\theta(\{x\})$ del suceso $\{x_1 = x\}$; aquí $\{x\}$ significa un conjunto compuesto por un solo punto x. También cabe señalar que, por ejemplo, la distribución normal Φ_{α,σ^2} y la distribución de Poisson son recíprocamente singulares. En vez de la medida de Lebesgue y la medida de cálculo también podríamos tomar otras medidas, por ejemplo, la distribución normal $\Phi_{0,1}$ y la distribución de Poisson Π_1, respectivamente. Sin embargo, en este caso las densidades $f_\theta(x)$ serán, evidentemente, otras. Proponemos que las halle el propio lector. Los ejemplos citados más arriba se referían al caso $\mathscr{X} = R$ o $\mathscr{X} = R^m$, $m > 1$. En un espacio de fase arbitrario $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}})$, la naturaleza de la medida μ puede ser más compleja.

La introducción de la condición $(A_\mu$ es cómoda, ante todo, por el hecho de que posteriormente nos permitirá examinar, desde un punto de vista único, dos tipos de distribuciones que son las más importantes en las aplicaciones: absolutamente continuas y discretas. Desde el punto de vista de la condición (A_μ), entre dichas distribuciones no hay ninguna diferencia cualitativa. Además, deja de ser importante la dimensión del espacio de fase $\mathscr{X}$

Convengamos en escribir

$$f(x) = g(x) \quad \text{c.d.} \ [\mu]$$

si existe un conjunto A, $\mu(A) = 0$ tal que $f(x) = g(x)$ para todos $x \notin A$. Es evidente que $f(x) = g(x)$ c.s. $[\mu]$ si y sólo si

$$\int (f(x) - g(x))^2 \mu(dx) = 0.$$

Lema 1. *Sean f y g dos densidades de probabilidad con respecto a la medida μ. Entonces*

$$\int f(x) \ln f(x) \mu(dx) \geqslant \int f(x) \ln g(x) \mu(dx), \tag{1}$$

si estas dos integrales son finitas. El signo de igualdad sólo es posible en el caso de $f = g$ c.d. $[\mu]$.

Aquí se vino al acuerdo de que las integrales en (1) sobre el conjunto A, en el que $f(x) = 0$, equivalen a cero para cualquier $g(x)$.

Demostración. Es necesario demostrar que

$$\int f(x) \ln \frac{g(x)}{f(x)} \mu(dx) \leqslant 0.$$

Como $\ln(1 + x) \leqslant x$ para todos $x \geqslant -1$, y el signo de igualdad sólo es

posible cuando $x = 0$, entonces

$$\ln \frac{g(x)}{f(x)} = \ln\left(1 + \left(\frac{g(x)}{f(x)} - 1\right)\right) \leqslant \frac{g(x)}{f(x)} - 1,$$

y el signo de igualdad aquí sólo es posible cuando $f(x) = g(x)$. Por eso

$$\int f(x) \ln \frac{g(x)}{f(x)} \mu(dx) \leqslant \int f(x)\left(\frac{g(x)}{f(x)} - 1\right)\mu(dx) =$$

$$= \int g(x)\mu(dx) - \int f(x)\mu(dx) = 0. \quad (2)$$

Si la relación $f = g$ c.d. $[\mu]$ no tiene lugar, es evidente que el signo de desigualdad en (2) será estricto. ⊲

Examinemos ahora la familia $\mathscr{P} = \{\mathbf{P}_\theta\}_{\theta\in\Theta}$ que satisface las condiciones (A_0), (A_μ) y la "distancia" $d(\mathbf{P}_\theta, \mathbf{Q})$ entre la distribución arbitraria $\mathbf{Q}$ y la distribución $\mathbf{P}_\theta \in \mathscr{P}$

$$d(\mathbf{P}_\theta, \mathbf{Q}) = -\int \ln f_\theta(x)\mathbf{Q}(dx). \quad (3)$$

Definamos la funcional $G(\mathbf{Q})$ como el valor de θ con el que se alcanza

$$\min_\theta d(\mathbf{P}_\theta, \mathbf{Q}) = d(\mathbf{P}_{G(\mathbf{Q})}, \mathbf{Q}).$$

Del lema 1 y la condición (A_0) se deduce que

$$-\int f_{\theta_0} \ln f_\theta \mu(dx) > -\int f_{\theta_0} \ln f_{\theta_0}\mu(dx),$$

$$d(\mathbf{P}_\theta, \mathbf{P}_{\theta_0}) > d(\mathbf{P}_{\theta_0}, \mathbf{P}_{\theta_0})$$

cuando $\theta \neq \theta_0$. Esto significa que

$$G(\mathbf{P}_{\theta_0}) = \theta_0. \quad (4)$$

Definición 1. Llámase *estimación de máxima verosimilitud* (*e.v.m.*) el valor de $\hat{\theta}^* = G(\mathbf{P}_n^*)$, o sea, el valor de θ con el que se alcanza

$$\max_\theta \int \ln f_\theta(x)\mathbf{P}_n^*(dx) = \max_\theta \frac{1}{n} \sum_{i=1}^{n} \ln f_\theta(x_i). \quad (5)$$

En lo sucesivo, el símbolo ˆ sobre la designación de la estimación corresponderá siempre a la e.v.m.

De la definición y de (4) se deduce que la *e.v.m. es una estimación de sustitución*. Esta también puede ser considerada como la estimación con arreglo al valor mínimo de la distancia (3). Esta distancia se halla íntimamente ligada a la distancia de Kullback—Leibler entre las distribuciones, la cual desempeña un papel especial en la estadística matemática y será examinada más tarde.

En la definición 1, la familia $\{\mathbf{P}_\theta\}$ se supone tal que $\hat{\theta}^*$ sea una magnitud aleatoria *.

En vista de que el valor máximo de cierta función puede alcanzarse en varios puntos, la e.v.m., hablando en general, no es única. El ejemplo respectivo será expuesto un poco más tarde.

La denominación de dicha estimación está relacionada con la siguiente interpretación importante de la expresión

$$\sum_{i=1}^{n} \ln f_\theta(x_i) = \ln \prod_{i=1}^{n} f_\theta(x_i),$$

presente en (5). Para facilitar la exposición examinemos primero el caso discreto cuando μ es la medida de cálculo. Entonces $\prod_{i=1}^{n} f_\theta(x_i)$ es la probabilidad de que aparezca el resultado $X = (x_1, \ldots, x_n)$. Por lo tanto, elegimos, en calidad de $\hat{\theta}^*$, el valor del parámetro que *maximiza esta probabilidad* (pues las funciones $\varphi(\theta) > 0$ y $\ln \varphi(\theta)$ alcanzan los valores extremos en los mismos puntos).

Una interpretación análoga también tiene lugar en el caso general. En virtud de la independencia de x_i tenemos, para los conjuntos $B = B_1 \times \ldots \times B_n$, $B_i \in \mathfrak{B}_{\mathscr{X}}$,

$$\mathbf{P}_\theta(X \in B) = \int_{B_1} f_\theta(x_1)\mu(dx_1) \ \ldots \int_{B_n} f_\theta(x_n)\mu(dx_n). \qquad (6)$$

Recordemos que x_i, a distinción de los elementos de la muestra x_i, designan las variables aleatorias, y el vector $(x_1, \ldots, x_n)$ se designa a través de x. Supongamos que μ^n es el producto directo múltiplo de n de las medidas μ, así que $\mu^n(dx) = \prod_{i=1}^{n} \mu(dx_i)$. Entonces (6) significa que

$$\mathbf{P}_\theta(X \in B) = \int_B \prod_{i=1}^{n} f_\theta(x_i)\mu^n(dx)$$

y, por consiguiente, la función $f_\theta(x) = \prod_{i=1}^{n} f_\theta(x_i)$ es la densidad de distribución del vector aleatorio X en $\mathscr{X}^n$ respecto a la medida μ^n,

$$\int_{\mathscr{X}^n} f_\theta(x)\mu^n(dx) = 1.$$

Ahora bien, $\prod_{i=1}^{n} f_\theta(x_i)\mu^n(dx)$ puede interpretarse (análogamente al caso

---

*) O sea, $\hat{\theta}^*$ realiza la aplicación medible de $(\mathscr{X}^n, \mathfrak{B}^n_{\mathscr{X}})$ en $(R^k, \mathfrak{B}^k)$.

discreto) como la probabilidad de que la muestra adquiera el valor del paralelepípedo formado por la intersección de las "franjas" $(x_i, x_i + dx_i)$, y la estimación de la máxima verosimilitud maximiza en θ esta probabilidad.

La función

$$f_\theta(X) = \prod_{i=1}^{n} f_\theta(x_i)$$

como función de θ se llama *función de verosimilitud*, y la función

$$L(X, \theta) = \ln f_\theta(X) = \sum_{i=1}^{n} l(x_i, \theta),$$

donde $l(x, \theta) = \ln f_\theta(x)$, se denomina *función logarítmica de verosimilitud*.

Esas mismas denominaciones de las funciones f y L también se utilizarán en el caso cuando como argumento, en vez de X, se halle el vector variable x. Ahora bien, la función de verosimilitud $f_\theta(x)$ es la función sobre $\mathscr{X}^n \times \Theta$ que, para cada $\theta \in \Theta$, constituye la densidad de la probabilidad respecto a la medida μ^n, así que la densidad $f_\theta x_1)$ en $\mathscr{X}$ también es la función de verosimilitud para el caso $n = 1$.

Por otro lado, $f_\theta(X)$, por ejemplo, en el caso $\mathscr{X} = R$, puede considerarse como la función de verosimilitud de una muestra de volumen 1 en el caso multidimensional, cuando $\mathscr{X} = R^m = R^n$.

Cabe señalar que la e.v.m. no depende absolutamente de la elección de la medida μ, puesto que, al sustituir μ por cualquier medida equivalente μ_1, la función de verosimilitud $f_\theta(x)$ cambiará sólo en el factor $\frac{d\mu^n}{d\mu_1^n}(x)$ que no depende de θ.

Las propiedades asintóticas de la e.v.m. podrían haber sido investigadas en el mismo camino que utilizamos al estudiar las estimaciones por el método de momentos. Precisamente allí hemos aprovechado el hecho de que las estimaciones conforme al método de momentos son estadísticas de tipo I. Esto nos permitió determinar directamente su conciliabilidad fuerte y su normalidad asintótica. Al cumplirse ciertas condiciones para $f_\theta(x)$, las *e.v.m. serán estadísticas de tipo* II, y esto también permite (véanse los teoremas de los §§ 1.5, 1.8) determinar su conciliabilidad y su normalidad asintótica. No obstante, a nosotros nos será más cómodo estudiar directamente las propiedades de las e.v.d. (véanse los §§ 23—27), ya que esto permite realizar la investigación de un modo más económico y completo.

Hallemos las funciones de verosimilitud y las e.v.m. para algunas distribuciones expuestas en el § 2. En cuanto a las funciones de verosimilitud suaves, la manera más fácil de hallar su valor máximo consiste en igualar a cero las primeras derivadas.

Ejemplo 1. La distribución normal de Φ_{α,σ^2} en $\mathscr{X} = R$ tiene una densidad de

$$\varphi_{\alpha,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\alpha)^2}{2\sigma^2}}, \quad -\infty < \alpha < \infty, \ \sigma > 0.$$

Suponiendo, en este caso, que $\theta = (\alpha, \sigma^2)$, obtenemos

$$f_\theta(x) = (2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \alpha)^2\right\},$$

$$L(X, \theta) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \alpha)^2.$$

En vista de que ln es una función monótona, como ya hemos señalado, f y L alcanzan su valor máximo con los mismos valores de θ. Tenemos

$$\frac{\partial L}{\partial \alpha} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \alpha),$$

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \alpha)^2.$$

Resolviendo, para el punto del valor máximo, el sistema de ecuaciones

$$\frac{\partial L}{\partial \alpha} = 0, \quad \frac{\partial L}{\partial \sigma} = 0,$$

obtenemos

$$\hat{\alpha}^* = \bar{x}, \quad (\hat{\sigma}^2)^* = S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Es fácil comprobar que en este punto realmente se alcanza el valor máximo de L.

Ejemplo 2. Examinemos la distribución Γ con densidad

$$\gamma_\alpha(x) = \frac{\alpha^\lambda}{G(\lambda)} x^{\lambda-1}e^{-\alpha x}, \quad x \geqslant 0, \ \alpha > 0,$$

en el caso cuando se conoce el parámetro λ. Tenemos

$$L(X, \alpha) = \lambda n \ln\alpha - n\ln\Gamma(\lambda) + (\lambda - 1)\sum_{i=1}^{n}\ln x_i - \alpha\sum_{i=1}^{n}x_i,$$

$$\frac{\partial L}{\partial \alpha} = \frac{\lambda n}{\alpha} - \bar{x}n, \quad \hat{\alpha}^* = \lambda/\bar{x}.$$

Ejemplo 3. Tenemos la distribución binomial $\mathbf{B}_p$. Aquí, para $X \in \mathbf{B}_p$ tenemos que $\mathbf{P}(x_i = 1) = p$, $\mathbf{P}(x_i = 0) = 1 - p$,

$$f_p(X) = p^\nu(1-p)^{n-\nu},$$

donde ν es el número de apariciones de 1 entre los elementos $x_1, \ldots, x_n$. Por lo tanto,

$$L(X, p) = \nu \ln p + (n - \nu) \ln(1 - p),$$

$$\frac{\partial L}{\partial p} = \frac{\nu}{p} - \frac{n - \nu}{1 - p}, \quad \hat{p}^* = \frac{\nu}{n}.$$

Proponemos al lector que procure, en forma de ejercicio, hallar las e.v.m. para todas las familias paramétricas expuestas en el § 2, y que las compare con las estimaciones según el método de momentos.

Ahora citaremos dos ejemplos de un tipo, algo diferente, cuando la función f_θ no es suave en θ y cuando no son vigentes los métodos de búsqueda de la e.v.m., relacionados con la derivación.

Ejemplo 4. Sea $X \in U_{\theta, 1+\theta}$ (distribución uniforme sobre $[\theta, 1 + \theta]$). Aquí

$$f_\theta(x) = \begin{cases} 1, & x \in [\theta, 1 + \theta], \\ 0, & x \notin [\theta, 1 + \theta], \end{cases}$$

$$f_\theta(X) = \begin{cases} 1, & \theta \leqslant x_{(1)} \leqslant x_{(n)} \leqslant 1 + \theta, \\ 0, & \text{de lo contrario,} \end{cases}$$

donde $x_{(1)} \leqslant \ldots \leqslant x_{(n)}$ es la serie variacional. En este ejemplo, la estimación de verosimilitud máxima no es única. En efecto, $f_\theta(X) = 1$ (o sea, al valor máximo) para todos los valores de θ que satisfagan las relaciones $x_{(n)} - 1 \leqslant \theta \leqslant x_1$. Como $x_{(n)} - x_{(1)} \leqslant 1$, tales θ existen siempre. Podemos tomar, en particular, $\hat{\theta}^* = x_{(1)}$ o bien $\hat{\theta}^* = x_{(n)} - 1$.

Ejemplo 5. Sea $X \in \mathbf{U}_{0,\theta}$. Aquí

$$f_\theta(x) = \begin{cases} \theta^{-1}, & x \in [0, \theta], \\ 0 \quad , & x \notin [0, \theta], \end{cases}$$

$$f_\theta(X) = \begin{cases} \theta^{-n} & \text{si } x_i \in [0, \theta] \text{ para todos } i = 1, 2, \ldots, n. \\ 0, & \text{de lo contrario.} \end{cases}$$

Para obtener la forma de función $f_\theta(X)$ como función de θ, escribamos la condición $x_i \in [0, \theta)$, $i = 1, \ldots, n$, en la forma equivalente $\theta \geqslant \text{máx}\, x_i = x_{(n)}$. Así pues, $f_\theta(X) = 0$ cuando $\theta \in [0, x_{(n)})$, y $f_\theta(X) = \theta^{-n}$ cuando

$\theta \in (x_{(n)}, \infty)$. El gráfico de esta función se muestra en la fig. 1. Aquí, al igual que en el ejemplo precedente, la función f_θ es discontinua. El valor máximo de f_θ se alcanza en el punto $\hat{\theta}^* = x_{(n)}$.

Análogamente el lector puede hallar la e.v.m. para un parámetro bidimensional desconocido (α, β) cuando $X \in U_{\alpha,\beta}$.

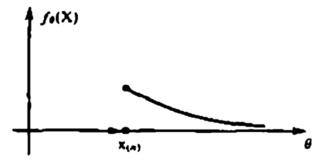


Fig. 1.

Si $f_\theta(x)$ es ilimitada y los puntos x_θ, en los que $f_\theta(x_\theta) = \infty$, dependen de θ, el método de verosimilitud máxima pierde en sumo grado su significado (aquí hemos venido al acuerdo de que $f_\theta(x_\theta) = \infty$ si $f_\theta(x) \to \infty$ cuando $x \downarrow x_\theta$ o cuando $x \uparrow x_\theta$). Esto se puede entender con más facilidad en el ejemplo del parámetro de desplazamiento cuando $f_\theta(x) = f(x - \theta)$, $f(x) > 0$, $f(0) = \infty$. Entonces $f_\theta(X) = \infty$ cuando $\theta = x_1, \ldots, \theta = x_n$ y, por consiguiente, $\hat{\theta}^*$ adquiere, por lo menos, n valores que coinciden con los elementos de la muestra. La esencia de tal efecto consiste en que en este caso los "saltos" de $f_\theta(X)$ no dan la posibilidad de juzgar acerca de la posición del máximo "verdadero" de $f_\theta(X)$, determinado por la influencia de toda la muestra (compárese esto con los §§ 24, 25). Para obtener tal parámetro sería necesario "amortiguar" de algún modo los saltos de $f_\theta(X)$.

Las estimaciones de verosimilitud máxima poseen la siguiente propiedad importante de invariancíón con respecto a la sustitución del parámetro.

Teorema 1. *Supongamos que $\beta(\theta)$ es la función que realiza la aplicación biunívoca del conjunto Θ sobre el conjunto B. Entonces, si θ^* es la e.v.m. segun la muestra X del parámetro θ, en este caso $\beta^* = \beta(\hat{\theta}^*)$* será la e.v.m. según la muestra X del parámetro $\beta = \beta(\theta)$ *para la familia paramétrica* $\{Q_\beta = P_{\theta(\beta)}\}_{\beta \in B}$, *donde $\theta(\beta)$ es la función inversa a $\beta(\theta)$.*

Omitimos la **demostración** del teorema, debido a su evidencia.

Debemos señalar que ya hemos utilizado implícitamente el teorema 1 en el ejemplo 1, donde en busca de la e.v.m para σ^2 hemos hallado el valor máximo de L por σ y luego hemos tomado $(\hat{\sigma}^2)^* = \hat{\sigma}^*)^2$.

Otro ejemplo de uso de este teorema es la determinación de la e.v.m. en el caso de $X \in L_{\alpha,\sigma^2}$, o sea, en el caso cuando la distribución de x_i es lognormal: $\ln x_i \in \Phi_{\alpha,\sigma^2}$. Para tales x_i la media a y la varianza d^2 son iguales respectivamente (véase el § 2):

$$a = \exp\{\alpha + \sigma^2/2\}, \quad d^2 = a^2(e^{\sigma^2} - 1).$$

Si designemos por $\hat{a}^*$ y $(\hat{d}^2)^*$ las e.v.m. para a y d^2, en virtud de la propiedad de invariación obtenemos, para la función $(a, d^2) = \beta(\alpha, \sigma^2)$ (véase el ejemplo 1),

$$\hat{a}^* = \exp\left\{\bar{y} + \frac{S_Y^2}{2}\right\}, \quad (\hat{d}^2)^* = (\hat{a}^*)^2\,(e^{S_Y^2} - 1),$$

donde $Y = (y_1, \ldots, y_n)$, $y_i = \ln x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$.

El cálculo aproximado de las e.v.m. en situaciones más complicadas se realiza en el § 26.

Para resumir este párrafo haremos la observación siguiente. Ya hemos dicho que la e.v.m. es una estimación de sustitución. No obstante, dicha e.v.m. también puede considerarse, en ciertas condiciones, como estimación del método generalizado de momentos. En efecto, supongamos que la función $f_\theta(x)$ es derivable respecto a θ y que es legítima la derivación respecto a esta variable bajo el signo integral en la igualdad

$$\int f_\theta(x)\mu(dx) = 1.$$

Entonces

$$0 = \int f_\theta'(x)\mu(dx) = \int_{\{f_\theta(x) \neq 0\}} \frac{f_\theta'(x)}{f_\theta(x)}\, f_\theta(x)\mu(dx) =$$

$$= \int_{\{f_\theta(x) \neq 0\}} l'(x, \theta) f_\theta(x)\mu(dx) = \mathbf{M}_\theta l'(x_1, \theta).$$

Ahora bien, si en (4.6) ponemos $g(x, \theta) = l'(x, \theta)$, para la estimación por el método generalizado de momentos obtenemos la ecuación

$$\int l'(x, \theta)\mathbf{P}_n^*(dx) = \int l'(x, \theta)\mathbf{P}_\theta(dx) = 0$$

o bien, que es lo mismo,

$$L'(X, \theta) = 0.$$

Esta es la ecuación para la e.v.m.

§ 7. Acerca de la comparación de las estimaciones

Hemos visto que existen muchos enfoques naturales de obtención de las estimaciones. Cabe preguntar: ¿cómo comparar entre sí diferentes estimaciones y qué estimaciones deben preferirse a otras? Destaquemos dos enfoques de comparación de las estimaciones: estándar (medio cuadrático o típico) y asintótico.

El primero de ellos se basa en la comparación de las desviaciones estándar. El segundo enfoque es aplicable solamente a las muestras de gran volu-

men, puesto que se funda en la comparación de las "dispersiones" de las distribuciones para $(\theta^* - \theta)\sqrt{n}$ en caso de grandes n. Como base para tal comparación sirve generalmente la forma de distribuciones límite para $\theta^* - \theta)\sqrt{n}$ cuando $n \to \infty$ (si éstas existen). Los teoremas límite respectivos nos dan las condiciones en las que la distribución $(\theta^* - \theta)\sqrt{n}$ para grandes n puede ser aproximada con ayuda de las distribuciones límite mencionadas.

En este párrafo se supone que las estimaciones se comparan en caso de una distribución desconocida cualquiera de la muestra **P**, pero registrada.

1. Enfoque estándar. Caso unidimensional. Este enfoque se utiliza para examinar las estimaciones con arreglo a la muestra X de cualquier volumen registrado (no obligatoriamente grande). Consiste en la comparación de las desviaciones típicas $\mathbf{M}(\theta^* - \theta)^2$.

Regla 1. Con arreglo al enfoque estándar, consideraremos que la estimación θ_1^* es mejor que la θ_2^* si

$$\mathbf{M}(\theta_1^* - \theta)^2 < \mathbf{M}(\theta_2^* - \theta)^2.$$

Está ampliamente difundida la idea de que el error estándar es la característica numérica más conveniente de la exactitud de una estimación, aunque desde muchos puntos de vista esta circunstancia es discutible: pues se puede comparar, digamos, las magnitudes $\mathbf{M}|\theta^* - \theta|$ que también describen los valores medios de las desviaciones de θ^* de θ.

La ventaja indudable de las características $\mathbf{M}(\theta^* - \theta)^2$ consiste en el hecho de que $(\theta^* - \theta)^2$ es la función analítica de la diferencia $\theta^* - \theta$. Esto hace más cómodos muchos estudios y permite aproximar, como veremos más tarde, los valores de $\mathbf{M}f(\theta^* - \theta)$ para las funciones suaves f.

A la par con la desviación estándar para la descripción de las propiedades de las estimaciones también se utiliza la magnitud de desplazamiento.

Definición 1. Se llama *desplazamiento* de la estimación θ^* la magnitud

$$b = \mathbf{M}\theta^* - \theta.$$

La estimación θ^*, para la cual $b = 0$, se denomina *no desplazada*.

La desviación estándar está relacionada con el desplazamiento y la varianza de la estimación por medio de la igualdad

$$\mathbf{M}(\theta^* - \theta)^2 = \mathbf{D}\theta^* + b^2,$$

así que para las estimaciones no desplazadas, la desviación estándar coincide con la varianza.

El carácter de no desplazamiento propiamente dicho es, evidentemente, una propiedad deseable de las estimaciones, puesto que significa que en la sucesión dad de estimaciones, el valor medio de éstas coincidirá con

el valor verdadero del parámetro. Si falta dicha propiedad, la estimación se llama *desplazada*.

Ejemplo 1. Examinemos las tres estimaciones siguientes para el valor medio $\theta = \mathbf{M}x_1$ de la distribución $\mathbf{P}$:

$$\theta_1^* = \bar{x}, \ \theta_2^* = \zeta^*, \ \theta_3^* = \frac{x_{(1)} + x_{(n)}}{2}, \tag{2}$$

donde ζ^* es la mediana muestral; $x_{(k)}$, $k = 1, \ldots, n$, los valores de la serie variacional, así que $\zeta^* = x_{((n+1)/2}$ si n es impar, y $\zeta^* = \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)})$ si n es par (para $n = 1, 2$ todas las tres estimaciones coinciden). Todas las estimaciones son no desplazadas si la distribución $\mathbf{P}$, de la que ha sido extraída la muestra, es simétrica con respecto a $\theta(\mathbf{P}((-\infty, \theta - t)) = \mathbf{P}((\theta + t, \infty))$ para cualquier $t \geqslant 0$). Esto se deduce del hecho de que la distribución de todas las tres estimaciones también será simétrica respecto a θ. Para $\bar{x}$, la afirmación sobre el no desplazamiento de $\mathbf{M}\bar{x} = \theta$ es evidente incluso sin la suposición acerca de la simetría.

Calculemos las desviaciones estándar de las estimaciones (2). Para simplificar la exposición nos limitaremos al caso de $\mathbf{P} = \mathbf{U}_{0,1}$, $n = 3$, para el cual las estimaciones (2) pasarán a

$$\theta_1^* = \bar{x}, \ \theta_2^* = x_{(2)}, \ \theta_3^* = \frac{x_{(1)} + x_{(3)}}{2}.$$

Tenemos

$$\mathbf{D}x_1 = \int_0^1 (x - 1/2)^2 dx = 1/12, \ \mathbf{M}(\theta_1^* - \theta)^2 = \mathbf{D}\bar{x} = Dx_1/3 = 1/36.$$

Luego, en virtud de la definición de la mediana (n es impar) $\{\zeta^* < x\} = \{F_n^*(x) > 1/2\}$ y, por lo tanto

$$\mathbf{P}(\zeta^* < x) = \mathbf{P}(F_n^*(x) > 1/2) = \sum_{k=(n+1)/2}^{n} \mathbf{P}(nF_n^*(x) = k). \tag{3}$$

Para $n = 3$,

$$\mathbf{P}(F_3^*(x) = 1) = \mathbf{P}\Big(\bigcap_{i=1}^{3} \{x_i < x\}\Big) = F^3(x),$$

$$\mathbf{P}(3F_3^*(x) = 2) = 3F^2(x)(1 - F(x)).$$

La probabilidad $\mathbf{P}(\zeta^* \in (u, u + du))$ se compone de las probabilidades de sucesos que tienen la forma $\{x_1 \in (u, u + du)\}\{x_2 < u\}\{x_3 > u\}$. Como en total son posibles 6 de estas combinaciones, $\mathbf{P}(\zeta^* \in (u, u + du)) =$

$= 6f(u)F(u)(1 - F(u))du$ y, por consiguiente, ζ^* tiene una densidad igual a (esto también resulta de (3))

$$6f(u)F(u)(1 - F(u)),$$

donde $F(u) = \int_{-\infty}^{u} f(t)dt = \mathbf{P}(x_1 < u)$. En el caso de $\mathbf{P} = \mathbf{U}_{0,1}$ esta densidad será igual a $6x(1 - x)$ cuando $x \in [0, 1]$, así que

$$\mathbf{M}(\zeta^*)^2 = \int_0^1 6x^3(1 - x)dx = 6\left(\frac{1}{4} - \frac{1}{5}\right) = \frac{3}{10},$$

$$\mathbf{D}\zeta^* = \mathbf{M}(\zeta^*)^2 - (\mathbf{M}\zeta^*)^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}.$$

Nos queda hallar la varianza de la estimación

$$\theta_3^* = \frac{x_{(1)} + x_{(3)}}{2}.$$

Razonando análogamente a la precedente, no es difícil convencerse de que la probabilidad $\mathbf{P}(x_{(1)} \in (u, u + du), x_{(3)} \in (v, v + dv))$, cuando $u < v$, es igual a $6f(u)f(v)\,(F(v) - F(v))du\,dv$. Por eso para $\mathbf{P} = \mathbf{U}_{0,1}$

$$\mathbf{M}(\theta_3^*)^2 = \int_0^1 \int_0^v \left(\frac{u + v}{2}\right)^2 6(v - u)du\,dv.$$

El valor de esta integral es igual a 11/40 (el lector puede realizar los cálculos individualmente), por lo tanto,

$$\mathbf{D}\theta_3^* = \mathbf{M}(\theta_3^*)^2 - (\mathbf{M}\theta_3^*)^2 = \frac{11}{40} - \frac{1}{4} = \frac{1}{40}.$$

Así pues, la estimación θ_3^* resulta la mejor. Para otros valores de n y otras distribuciones $\mathbf{P}$, la situación puede ser otra. Veremos, por ejemplo, que cuando $\mathbf{P} = \Phi_{\alpha,\sigma^2}$, la mejor estimación para α será $\theta_1^* = \bar{x}$.

Ejemplo 2. *Estimaciones no desplazadas de la varianza.* Examinemos la estimación para la varianza

$$S^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 - (\bar{x})^2,$$

así como la estimación

$$S_1^2 = \frac{1}{n}\sum(x_i - \mathbf{M}x_i)^2 = \frac{1}{n}\sum x_i^2 + (\mathbf{M}x_1)^2 - 2\bar{x}\mathbf{M}x_1$$

(ambas según el principio de sustitución) en el caso cuando se conoce $\mathbf{M}x_1$.

La estimación S_1^2 no está, evidentemente, desplazada. Al mismo tiempo

$$S^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 = \frac{1}{n}\sum (x_i - \bar{x} \pm Mx_1)^2 =$$

$$= \frac{1}{n}\sum (x_i - Mx_1)^2 - (\bar{x} - Mx_1)^2 = S_1^2 - (\bar{x} - Mx_1)^2 < S_1^2.$$

Ahora bien, la estimación S^2 está desplazada,

$$MS^2 = Dx_1 - D\bar{x} = \left(1 - \frac{1}{n}\right)Dx_1.$$

Esta relación muestra que también podemos examinar, en caso de Mx_1 desconocida, la estimación de la varianza igual a

$$S_0^2 = \frac{n}{n-1} S^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2, \quad MS_0^2 = Dx_1.$$

Pasemos ahora al enfoque asintótico del problema de comparación de las estimaciones. En este caso la regla para la preferencia de las estimaciones se elige unívocamente.

2. Enfoque asintótico. Caso unidimensional. Supongamos que se han dado dos estimaciones θ_1^* y θ_2^* tales que

$$\frac{(\theta_1^* - \theta)\sqrt{n}}{\sigma_1} \in Q, \quad \frac{(\theta_2^* - \theta)\sqrt{n}}{\sigma_2} \in Q, \tag{4}$$

donde Q es cierta ley de distribución límite, la misma que para θ_1^* y θ_2^*, y $\sigma_2 > \sigma_1$. Entonces, para grandes valores de n, las distribuciones $(\theta_i^* - \theta)\sqrt{n}/\sigma_i$, $i = 1, 2$ serán próximas a Q, e indudablemente que la "dispersión" de θ_2^* alrededor de θ será mayor que la "dispersión" de θ_1^* y debemos preferir θ_1^*.

Ahora bien, la esencia del enfoque asintótico consiste en la comparación de las distribuciones límites de las estimaciones.

Ya hemos visto y también nos convenceremos de ello ulteriormente, que muchas estimaciones aparecidas de un modo natural, icluyendo las óptimas (de lo cual hablaremos posteriormente), son asintóticamente normales, o sea, para ellas es válida (4) cuando $Q = \Phi_{0,1}$. Esto nos permite enunciar la siguiente regla natural de comparación de las estimaciones a.n.

Supongamos que se dan dos estimaciones a.n. θ_1^* y θ_2^* con los coeficientes σ_1^2 y σ_2^2, respectivamente.

Regla 2. La estimación θ_1^* debe ser mejor que θ_2^* si $\sigma_1^2 < \sigma_2^2$.

En lo sucesivo, al utilizar estas y otras reglas, a la par con el término "mejor" también haremos uso, donde sea necesario, de las palabras "no peor", "peor", "no mejor" que corresponderán a los signos de desigualdad $\leqslant, >, \geqslant$ entre σ_1^2 y σ_2^2 (o bien entre $M(\theta_1^* - \theta)^2$ y $M(\theta_2^* - \theta)^2$ en (1)). Si

$\sigma_1^2 = \sigma_2^2$, diremos que estas estimaciones son asintóticamente equivalentes. El acuerdo propuesto es natural, y en las definiciones ulteriores no lo mencionaremos cada vez y sólo nos limitaremos a difinir la relación "mejor" o las relaciones semejantes a ésta.

Es preciso señalar que en la clase de estimaciones a.n., la minimalidad de la dispersión de θ^* quiere decir que la magnitud

$$\lim_{n\to\infty} \mathbf{P}(|\theta^* - \theta| < u/\sqrt{n})$$

será máxima para cada u. Esta circunstancia hace indiscutible la regla indicada para la comparación de las estimaciones a.n.

El enfoque asintótico, a pesar de su naturalidad, tiene una desventaja considerable: *sólo es aplicable a las estimaciones de gran volumen y únicamente en la clase de estimaciones a.n.*

Los dos enfoques señalados son, en cierto sentido, próximos uno a otro: en ambos casos el hecho se reduce a la comparación de las varianzas o de las magnitudes próximas a ellas. Por supuesto que la magnitud σ_1^2/n en (4), cuando $\mathbf{Q} = \Phi_{0,1}$, puede distinguirse considerablemente de $\mathbf{M}(\theta^* - \theta)^2$. Sin embargo, los ejemplos que ilustran este hecho (proponemos al lector que los construya él mismo) tienen, por lo común, carácter artificial.

La exposición ulterior de este capítulo está relacionada, en mucho, con la construcción de las estimaciones, óptimas para cada uno de los dos enfoques introducidos.

Ejemplo 3. Sea $X \in \Gamma_{\alpha,1}$. En el ejemplo 1 del § 4 hemos mostrado que ambas estimaciones

$$\alpha_1^* = (\bar{x})^{-1} \text{ y } \alpha_2^* = \left(\frac{1}{2n}\sum x_i^2\right)^{-1/2}$$

son estimaciones conforme al método de momentos. Además, α_1^* también es e.v.m. Luego hemos determinado que ambas estimaciones son asintóticamente naturales, con coeficientes α^2 y $\frac{5}{4}\alpha^2$, respectivamente, y por lo tanto, la estimación α_1^* es mejor que la α_2^* desde el punto de vista del enfoque asintótico. Ese mismo resultado, para $n \geqslant 2$, se obtiene cuando se trata del enfoque estándar.

Ahora citaremos un ejemplo que muestra que según las propiedades de la distribución, una misma estimación puede ser mejor o peor que alguna otra estimación registrada.

Ejemplo 4. Examinemos el problema de la estimación $\theta = \mathbf{M}x_1$ si se sabe que $X \in \mathbf{P}$ y que la distribución $\mathbf{P}$ es simétrica respecto al punto

θ (compárese con el ejemplo 1). En este caso la mediana de la distribución ζ coincide con θ. Examinemos también dos estimaciones para θ (ambas según el principio de sustitución): la media $\theta_1^* = \bar{x}$ y la mediana muestral $\theta_2^* = \zeta^*$. Supongamos, para precisar, que n es impar. Del corolario 2.2.1, cuando $k = (n + 1)/2$, se deduce que si la función de distribución F es continuamente derivable en el punto $\theta = \zeta$, entonces

$$(\zeta^* - \zeta)\sqrt{n} \Rightarrow \frac{\xi}{2f(\theta)}, \quad \xi \in \Phi_{0,1}, \quad f(x) = F'(x).$$

Con otras palabras, en este caso ζ^* es la estimación a.n. con coeficiente $\sigma_2^2 = 1/(4f^2(\zeta))$.

Por otro lado, la estimación a.n. de $\bar{x}$ tiene por coeficiente $\sigma_1^2 = \mathbf{D}x_1$. Ahora bien, si

$$\int (x - \zeta)^2 dF(x) < \frac{1}{4f^2(\zeta)},$$

debemos preferir la estimación $\bar{x}$. Si el signo de desigualdad es inverso, entonces debemos preferir ζ^*. Cabe señalar que los números $\int (x - \zeta)^2 dF(x)$ y $f(\zeta)$ son características de distribución muy poco relacionadas entre sí.

Examinemos un importante caso particular, cuando estimamos el parámetro α por la muestra $X \in \Phi_{\alpha,\sigma^2}$. En este caso $f(\alpha) = f(\zeta) = \dfrac{1}{\sigma\sqrt{2\pi}}$, así que

$$\sigma_2^2 = \frac{\pi}{2}\sigma^2 > \sigma^2 = \sigma_1^2.$$

Esto significa que en esta situación, la estadística $\bar{x}$ es mejor que la ζ^*. Sin embargo, como hemos visto, no es difícil construir el ejemplo de la distribución para la cual será preferible la estadística ζ^*.

El ejemplo de la mediana también es muy aleccionador en otro sentido. El mismo muestra que la velocidad de disminución del grado de dispersión de $\zeta^* - \zeta$ puede ser cualquiera. Para cerciorarse de esto, basta con recurrir a la observaicón 2.2.1. En condiciones de dicha observación, como factor normalizador que asegura la convergencia de $\zeta^* - \zeta$ hacia la distribución límite sirve la magnitud $n^{1/(2\gamma)}$, donde γ es un número no negativo cualquiera (véase (2.12)). El factor $\sqrt{n}$ corresponde solamente a las distribuciones suaves.

Ahora presentaremos un experimento real con la muestra de volumen $n = 101$ de la población normal $\Phi_{0,1}$ y veremos *) cómo los valores de $\bar{x}$

*) La muestra X ha sido construida con ayuda de los números aleatorios tomados de las tablas [8] (se han utilizado los primeros 101 números en la página).

y ζ^* aproximan el 0 cuando $n = 11, 21, 51, 101$. Los datos obtenidos se ofrecen en la tabla siguiente:

n	11	21	51	101
$\bar{x}$	−0,283	−0,254	−0,148	−0,072
ζ^*	−0,291	−0,292	−0,078	−0,044

En este ejemplo, la estimación ζ^* para $n = 51, 101$ se comporta mejor, lo cual es resultado de la desviación aleatoria. Para convencerse de la ventaja de $\bar{x}$ sería necesario realizar muchos experimentos de este tipo.

Veamos ahora que aspecto tienen los dos enfoques (anteriormente enunciados) de la comparación de las estimaciones en el caso multidimensional, cuando θ es el vector $(\theta_1, \ldots, \theta_k)$.

3. Enfoques estándar y asintótico en el caso multidimensional. Como antes, utilizaremos el enfoque asintótico sólo en la clase de estimaciones a.n. En este caso el hecho se reduce por completo a la comparación de las distribuciones normales multidimensionales (distribuciones límites para $(\theta^* - \theta)\sqrt{n}$) que se describen totalmente por medio de la matriz de segundos momentos σ^2 (véase, por ejemplo, el teorema 3.2A).

Si se examina el enfoque estándar de la comparación de las distribuciones exactas de θ^*, también todo se reduce a la posibilidad de comparar dos distribuciones en R^k, basándose en el conocimiento de los momentos $(\theta^* - \theta)$ de segundo orden. Ahora bien, en ambos casos debemos saber comparar, según el "grado de dispersión", las matrices de los momentos de segundo orden.

Examinemos los métodos de comparación más naturales. Supongamos que $\mathbf{Q}_1$ y $\mathbf{Q}_2$ son dos distribuciones aleatorias en R^4. Designemos por ξ_1 y ξ_2 cualesquiera vectores aleatorios que poseen estas distribuciones: $\xi_i \in \in \mathbf{Q}_i$.

Definición 2. Diremos que la dispersión estándar de la distribución $\mathbf{Q}_1$ alrededor del punto $\alpha \in R^k$ no es mayor que la dispersión $\mathbf{Q}_2$ si para todo vector $a = (a_1, \ldots, a_k)$,

$$\mathbf{M}(\xi_1 - \alpha, a)^2 \leqslant \mathbf{M}(\xi_2 - \alpha, a)^2, \tag{5}$$

donde $(x, a) = \sum_{i=1}^{b} x_i a_i$ es el producto escalar.

Diremos que la dispersión para $\mathbf{Q}_1$ es menor que para $\mathbf{Q}_2$ si en (5) tiene lugar el signo de desigualdad estricta al menos para un a.

Si $\alpha = \mathbf{M}\xi_1 = \mathbf{M}\xi_2$, la igualdad (5) significa que por cualquier dirección de a la varianza de la distribución $\mathbf{Q}_1$ (o sea, la varianza de la proyección de ξ_1 sobre a) no supera la magnitud igual para $\mathbf{Q}_2$.

Si $d_l^2 = \|d_{ij}^{(l)}\|$ es la matriz de segundos momentos de $\mathbf{Q}_l$, $l = 1, 2$, entonces, abriendo paréntesis en (5) para $\alpha = 0$, obtenemos, para todos $a_1, \ldots, a_k$

$$\sum_{i,j=1}^{k} d_{ij}^{(1)} a_i a_j \leqslant \sum_{i,j=1}^{k} d_{ij}^{(2)} a_i a_j. \tag{6}$$

En el lenguaje de las matrices designaremos esta relación por

$$d_1^2 \leqslant d_2^2, \tag{7}$$

que significa la definición no negativa de la matriz $d_2^2 - d_1^2$.

Ahora bien, la dispersión estándar de $\mathbf{Q}_1$ alrededor del cero no supera tal dispersión para $\mathbf{Q}_2$ si y sólo si para las matrices de los momentos de segundo orden tienen lugar las desigualdades (6) y (7).

Las reglas de preferencia de las estimaciones en el caso multidimensional pueden enunciarse del modo siguiente.

Enfoque estándar: la estimación θ_1^* es mejor que la θ_2^* si la dispersión estándar de θ_1^* alrededor del punto θ es menor que la misma magnitud para θ_2^*.

Si d_l^2 es la matriz de segundos momentos $\theta_l^* - \theta$, la afirmación que dice que "la estimación θ_1^* es mejor que la θ_2^*" significa que $d_1^2 < d_2^2$.

Enfoque asintótico: la estimación θ_1^* es mejor que la θ_2^* si la dispersión estándar cerca del cero de la distribución límite para $(\theta_1^* - \theta)\sqrt{n}$ es menor que la misma magnitud para $(\theta_2^* - \theta)\sqrt{n}$.

En otros términos, si $(\theta_l^* - \theta)\sqrt{n} \Subset \Phi_{0,\sigma_l^2}$, entonces la afirmación de que "la estimación θ_1^* es mejor que la θ_2^*" quiere decir que $\sigma_1^2 < \sigma_2^2$.

Se puede mostrar que si θ_1^* y θ_2^* son dos estimaciones a.n. y θ_1^* es mejor que θ_2^*, entonces

$$\lim_{n\to\infty} \mathbf{P}((\theta_1^* - \theta)\sqrt{n} \in B) > \lim_{n\to\infty} \mathbf{P}((\theta_2^* - \theta)\sqrt{n} \in B) \tag{8}$$

para cualquier elipsoide central *) B.

Vemos que en ambos casos la comparación de las estimaciones se reduce al establecimiento de las igualdades para las matrices de los momentos de segundo orden. Cierta diferencia consiste en que en el primer caso los momentos no son obligatoriamente centrales.

Establezcamos ahora ciertas relaciones equivalentes a (6), (7).

---

*) Para abreviar convengamos en llamar *elipsoide* en R^k el dominio $\sum_{i,j=1}^{k} d_{ij}x_ix_j \leqslant c$, y *elipse*, la superficie $\sum_{i,j=1}^{k} d_{ij}x_ix_j = c$.

Pongamos

$$v(\theta^*) = \mathbf{M}(\theta^* - \theta)V(\theta^* - \theta)^T$$

y designemos por $\mathfrak{B}_+$ el conjunto de todas las matrices $V = |v_{ij}|$ definidas no negativamente. Si $|d_{ij}|$ es la matriz de segundos momentos $\theta^* - \theta$, entonces, evidentemente, $v(\theta^*) = \sum v_{ij}d_{ij}$.

Lema 1. *$d_1^2 \leqslant d_2^2$ si y sólo si $v(\theta_1^*) \leqslant v(\theta_2^*)$ para cualesquier $V \in \mathfrak{B}_+$.*

Demostración. En una dirección la afirmación es evidente, ya que la matriz $V_a = |a_i a_j| \in \mathfrak{B}_+$, y para tal matriz,

$$v_a(\theta_l^*) = \mathbf{M}(\theta_l^* - \theta)V_a(\theta_l^* - \theta)^T = \sum a_i a_j d_{ij}^{(l)}$$

(véase (6)).

Para demostrar la afirmación en dirección contraria, señalemos que el orden parcial basado en las desigualdades (5) es invariante respecto a los ejes de revolución de las coordenadas. Es decir, si C es la matriz de transformación ortogonal y θ_1^* es mejor que θ_2^* para el parámetro θ, entonces θ_1^*C es mejor que θ_2^*C para el parámetro θC. Esto se deduce de las igualdades

$$(\theta_l^*C - \theta C,\ a) = ((\theta_l^* - \theta)C,\ a) = (\theta_l^* - \theta,\ aC^T)$$

y de la definición 2.

Supongamos ahora que $d_1^2 < d_2^2$, o sea,

$$\sum d_{ij}^{(1)}a_i a_j < \sum d_{ij}^{(2)}a_i a_j. \tag{9}$$

Esto quiere decir que $v(\theta_1^*) < v(\theta_2^*)$ para las matrices V que tienen la forma $V_a = |a_i a_j|$ y, por lo tanto, también para las matrices diagonales $V_{\text{diag}} \in \mathfrak{B}_+$, puesto que estas últimas son representables en forma de la suma de k matrices que tienen la forma V_a. Supongamos ahora que V es una matriz arbitraria de $\mathfrak{B}_+$ y C es una transformación ortogonal tal que $C^T VC = V_{\text{diag}}$. Entonces

$$v(\theta_1^*) = \mathbf{M}(\theta_1^* - \theta)V(\theta_1^* - \theta)^T = \mathbf{M}(\theta_1^* - \theta)CV_{\text{diag}}C^T(\theta_1^* - \theta)^T.$$

De las dos observaciones hechas anteriormente y de (9) se deduce que el segundo miembro de esta igualdad es menor que

$$\mathbf{M}(\theta_2^* - \theta)CV_{\text{diag}}C^T(\theta_2^* - \theta)^T = \mathbf{M}(\theta_2^* - \theta)V(\theta_2^* - \theta)^T = v(\theta_2^*). \ \triangleleft$$

Existe también otro método de comparar la dispersión (véase [37]) que, sin embargo, supone que ambas distribuciones $\mathbf{Q}_1$ y $\mathbf{Q}_2$ no están degeneradas en R^k y tienen una media nula. En este caso las matrices de los segundos momentos centrales d_l^2 quedarán definidas positivamente y para ellas existen las inversas $A_l = (d_l^2)^{-1}$.

Supongamos que d^2 es la matriz de segundos momentos de la distribución $\mathbf{Q}$, y que $A = (d^2)^{-1}$.

Definición 3. Se llama *elipsoide de dispersión de la distribución* **Q** el elipsoide

$$tAt^T \leqslant k + 2$$

que entre todos los elipsoides se destaca unívocamente por su propiedad siguiente: si se examina la distribución uniforme **U** (o sea, la distribución en R^k con densidad constante dentro del elipsoide y con densidad nula fuera de éste), en este elipsoide, los primeros y segundos momentos de **Q** y de **U** coinciden (véase [25], p.333).

Lema 2. *Supongamos que las matrices d_l^2, $l = 1, 2$, no han sido degeneradas. La dispersión estándar de $\mathbf{Q}_1$ alrededor del cero no es mayor que la dispersión de $\mathbf{Q}_2$ si y sólo si el elipsoide de dispersión para $\mathbf{Q}_1$ se encuentra en el elipsoide para $\mathbf{Q}_2$.*

Demostración. Supongamos que la elipse $tA_1t^T = 1$ se encuentra en el interior de $tA_2t^T = 1$. Como es sabido, existe la transformación lineal no degenerada $t = uL$ que transfiere la elipse $1A_1t^T = 1$ a la esfera unitaria S_1, y la elipse $tA_2t^T = 1$, a la elipse S_2 con los ejes principales en dirección de los ejes de coordenadas. Esto quiere decir que $\tilde{A}_1 = LA_1L^T = E$ (matriz unidad), $\tilde{A}_2 = LA_2L^T = \operatorname{diag}(\lambda_1^2, \ldots, \lambda_k^2)$, $0 < \lambda_j^2 \leqslant 1$, $j = 1, \ldots, k$. Como $\tilde{A}_1^{-1} = E$, $\tilde{A}_2^{-1} = \operatorname{diag}(\lambda_1^{-2}, \ldots, \lambda_k^{-2})$, la elipse $t\tilde{A}_2^{-1}t^T = 1$ será una inversión respecto a la esfera unitaria S_1 de la elipse S_2 y, por consiguiente, se encontrará en S_1. Como $\tilde{A}_2^{-1} = (L^T)^{-1}A_2L^{-1}$, entonces, efectuando la transformación "inversa" $u = tL^T$, obtenemos que la elipse $tA_1^{-1}t^T = td_1^2t = 1$ se halla fuera de $tA_2^{-1}t^T = td_2^2t^T = 1$. Evidentemente, la misma relación es válida para las elipses $td_1^2t^T = c$ y $td_2^2t^T = c$. Pero esto significa que la igualdad $td_1^2t^T = c$ conduce a $td_1^2t^T = c \leqslant td_2^2t^T$. La afirmación en dirección contraria se muestra exactamente de la misma manera. ◅

Ahora es importante señalar que, a distinción del caso unidimensional, la comparación de las dispersiones con ayuda de las matrices de segundos momentos sólo establece el orden parcial en el conjunto de todas las distribuciones. Por ejemplo, las matrices $d_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ y $d_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ no son ni mejor ni peor una que otra, ya que para el vector $a = (1, 0)$, (6) es válida, y para el vector $a = (0, 1)$, la desigualdad será inversa. Esto constituye una incomodidad considerable del orden introducido, aunque éste, como tal, no suscita dudas.

Podemos hacer muchas estimaciones (o muchas distribuciones) bien ordenadas, si comparamos, digamos, $\mathbf{M}|\theta^* - \theta|^2$, donde $|\cdot|$ es la norma euclídea en R^k, así que

$$\mathbf{M}|\theta^* - \theta^2| = \mathbf{M}\sum_{i=1}^{k}(\theta_i^* - \theta_i)^2. \tag{10}$$

Tal método de ordenación ya es discutible, puesto que en distintas circunstancias, la precisión en diversas direcciones puede apreciarse de modo diferente. Para considerar de algún modo esta circunstancia, se puede, en calidad de generalización, tener en cuenta la medida de exactitud

$$\nu(\theta^*) = \mathbf{M}(\theta^* - \theta) V(\theta^* - \theta)^T,$$

donde V es la matriz definida no negativamente (el caso (10) corresponde a $V = E$).

Del lema 1 se deduce que si la dispersión de θ_1^* alrededor de θ es menor que la dispersión de θ_2^*, entonces $\nu(\theta_1^*) < \nu(\theta_2^*)$. El caso inverso, hablando en general, es incorrecto: el cumplimiento de la desigualdad $\nu(\theta_1^*) < \nu(\theta_2^*)$ para una matriz cualquiera V (el orden completo propuesto más arriba se basa en una matriz registrada) no significa aún que la dispersión de θ_1^* alrededor de θ es menor que la dispersión de θ_2^*.

Pasemos ahora a examinar un importante caso *paramétrico*, cuando se estiman los parámetros desconocidos de las distribuciones de familias paramétricas.

§ 8. Comparación de las estimaciones en el caso paramétrico. Estimaciones eficientes

En el párrafo precedente hemos destacado dos enfoques (estándar y asintótico) de la comparación de la calidad de las estimaciones. Introduzcamos ahora algunos conceptos relacionados con estos enfoques en el caso paramétrico, cuando la distribución de la muestra X pertenece a cierta familia $\mathscr{P} = \{\mathbf{P}_\theta\}$. Al igual que antes, con los símbolos $\mathbf{M}_\theta$ y D_θ designamos la esperanza matemática y la varianza de la distribución $\mathbf{P}_\theta$.

1. Caso unidimensional. Recordemos que de acuerdo con el enfoque estándar debemos decir qu θ_1^* es mejor que θ_2^* si

$$d_1^2(\theta) = \mathbf{M}_\theta(\theta_1^* - \theta)^2 < \mathbf{M}_\theta(\theta_2^* - \theta)^2 = d_2^2(\theta). \tag{1}$$

Pero en el caso paramétrico, $d_l^2(\theta)$, $l = 1, 2$, son las funciones de θ y debemos decir "θ_1^* es mejor que θ_2^* *en el punto* θ" si $d_1(\theta) < d_2(\theta)$.

Análogamente sucede al utilizar el enfoque asintótico cuando se comparan las estimaciones a.n. para grandes volúmenes de la muestra n, confrontando sus distribuciones límites. La estimación θ_1^* se considera mejor que la θ_2^* en el punto θ, si en las relaciones

$$(\theta_l^* - \theta)\sqrt{n} \in \Phi_{0,\, \sigma_l^2(\theta)}, \quad l = 1, 2, \tag{2}$$

es justa $\sigma_1(\theta) < \sigma_2(\theta)$*).

---

*) Ya hemos señalado que en la amplia clase de casos $d_l^2(\theta) = n^{-1}\sigma_l^2 + o(n^{-1})$. Sin embargo, esto no se deduce de las definiciones de los números $d_l^2(\theta)$ y $\sigma_l^2(\theta)$.

Ahora bien, en ambos casos el problema de comparación de las estimaciones conduce al asunto de *comparación de las funciones*, digamos, $d_i(\theta)$, $\theta = \Theta$. Este conjunto no está ordenado, y en la clase de todas las estimaciones es posible introducir un orden parcial del modo siguiente.

Regla 1. La estimación θ_1^* es mejor que la θ_2^* si $d_1(\theta) \leqslant d_2(\theta)$ (o, respectivamente, $\sigma_1(\theta) \leqslant \sigma_2(\theta)$) para todos $\theta \in \Theta$ y al menos para un θ se cumple la desigualdad estricta $d_1(\theta) < d_2(\theta)$.

Si la estimación θ^* es tal que para ella existe la estimación θ_1^* que es mejor que θ^*, en estos casos se dice que θ^* es una *estimación inadmisible.*

Expongamos primeramente el enfoque estándar en el caso unidimensional y examinemos las posibilidades aquí existentes de comparar las estimaciones. Conviene señalar, ante todo, que desde el punto de vista de la definición citada no existe, hablando en general, la mejor estimación. O sea, no existe una estimación θ^* tal que para toda otra estimación θ_1^* sea válida la desigualdad $d(\theta) \leqslant d_1(\theta)$, donde $d_1(\theta)$ está definida en (1), y $d(\theta)$ corresponde a θ^*.

En efecto, si se toma la estimación $\theta_1^* = \theta_1 = \text{const} \in \Theta$, entonces $d_1^2(\theta) = \mathbf{M}_\theta(\theta_1^* - \theta)^2 = 0$ cuando $\theta = \theta_1$ y para la mejor estimación θ^* (si tal estimación existiera) se cumplirá $d^2(\theta_1) = \mathbf{M}_{\theta_1}(\theta^* - \theta_1)^2 = 0$. Como θ_1 es arbitrario, $d^2(\theta) \equiv 0$. Pero esto es posible únicamente en el caso "degenerado", cuando las observaciones determinan unívocamente el valor del parámetro θ. Por ejemplo, cuando $X \in \mathbf{I}_\theta$ o bien $X \in \mathrm{U}_{\theta,\theta+1}$ y $\Theta = \{1, 2, \dots\}$.

Ahora bien, la envolvente inferior de todas las funciones $d^2(\theta)$ es igual a cero, pero en el caso "no degenerado" esta función no se realiza para ninguna función θ^*.

El problema puede ser más interesante si se buscan las mejores estimaciones θ^* en unas u otras subclases de estimaciones que se eligen de un modo suficientemente racional. Uno de los métodos posibles de destacar tales subclases consiste en registrar el desplazamiento $b(\theta)$.

Definición 1. La estimación $\theta_0^* \in K$ se denomina *eficiente en la clase* K si para cualquier otra estimación $\theta^* \in k$ $\mathbf{M}_\theta(\theta_0^* - \theta)^2 \leqslant \mathbf{M}_\theta(\theta^* - \theta)^2$ cuando todos $\theta \in \Theta$.

La clase K_0 de las estimaciones no desplazadas desempeña un papel especial, o sea, la clase de las estimaciones para las cuales $b(\theta) \equiv 0$.

Las estimaciones eficientes en la clase $K_0 = \{\theta^* : \mathbf{M}_\theta\theta^* = \theta\}$ de estimaciones no desplazadas se llaman simplemente *eficientes.* De suerte que las estimaciones eficientes no son sino estimaciones no desplazadas con varianza mínima.

Como ya hemos señalado, la propiedad de carácter no desplazado es, como tal, indudablemente deseable, ya que significa la falta del error sistemático al utilizar la estimación.

La cuestión acerca de la existencia de las estimaciones con el desplaza-

miento dado $b(\theta)$ (en particular, de las estimaciones no desplazadas) se reduce a la resolubilidad de la ecuación integral con respecto a $g(x)$:

$$\int g(x)\mathbf{P}_\theta(X \in dx) = \theta + b(\theta), \tag{3}$$

donde $g(X) = \theta^*$; el primer miembro de esta ecuación es $\mathbf{M}_\theta\theta^*$.

Si está cumplida la condición (A_μ) y $f_\theta(x) = \prod_{i=1}^{n} f_\theta(x_i)$ es la función de verosimilitud, la ecuación toma la forma

$$\int g(x)f_\theta(x)\mu^n(dx) = \theta + b(\theta). \tag{4}$$

Cabe señalar que la solución (4) para $b(\theta)$ dada no siempre existe ni mucho menos y, en particular, no para *todas las familas* $\{\mathbf{P}_\theta\}$ *existen las estimaciones no desplazadas del parámetro* θ. Examinemos, por ejemplo, el esquema de Bernoulli con un parámetro desconocido p (la probabilidad del caso es $\{x_1 = 1\}$) y supongamos que nos hace falta estimar el parámetro $\theta = \varphi(p)$, donde φ es una función dada. Entonces la ecuación (4) para la estimación no desplazada tiene la forma

$$\sum g(x)f_\theta(x) = \theta$$

o bien, que es lo mismo,

$$\sum_{k=0}^{n} G(k)p^k(1-p)^{n-k} = \varphi(p), \tag{5}$$

donde $G(k) = \sum_{x \in A_k} g(x)$ y A_k es el conjunto de puntos x cuyas k coordenadas son iguales a 1. Pero el primer miembro de (5) es el polinomio de p de grado n. Esto significa que la ecuación (5) sólo puede ser resuelta si $\varphi(p)$ es un polinomio de grado no mayor de n.

Examinemos ahora la clase K_b de estimaciones con desplazamiento registrado $b(\theta)$ y supongamos que existe una estimación que es eficiente en K_b.

Teorema 1. *La estimación eficiente en* K_b *es única con una exactitud de hasta los valores sobre el conjunto* $A \subset \mathscr{X}^n$ *para el cual* $\mathbf{P}_\theta(A) = 0$ *cuando todos* $\theta \in \Theta$.

Demostración. Sean θ_0^*, θ_1^* dos estimaciones eficientes en K_b. Designemos

$$D = \mathbf{D}_\theta\theta_i^*,\ \Delta_i = \theta_i^* - \theta,\ \theta^* = \frac{\theta_0^* + \theta_1^*}{2},\ i = 0,\ 1.$$

Como

$$\begin{gathered}\left(\frac{\Delta_0 + \Delta_1}{2}\right)^2 + \left(\frac{\Delta_0 - \Delta_1}{2}\right)^2 = \frac{\Delta_0^2 + \Delta_1^2}{2} \\ \frac{\Delta_0 + \Delta_1}{2} = \theta^* - \theta,\ \Delta_0 - \Delta_1 = \theta_0^* - \theta_1^*,\end{gathered} \tag{6}$$

entonces

$$\mathbf{M}_\theta(\theta^* - \theta)^2 + \frac{1}{4}\mathbf{M}_\theta(\theta_0^* - \theta_1^*)^2 = D + b^2(\theta). \tag{7}$$

Pero $\theta^* \in K_b$ y, por lo tanto, $\mathbf{M}_\theta(\theta^* - \theta)^2 \geqslant D + b^2(\theta)$. En este caso, de (7) se deduce que

$$\mathbf{M}_\theta(\theta_0^* - \theta_1^*)^2 \leqslant 0,$$

$\theta_1^* = \theta_0^*$ c.s.[*]. ◁

El análisis realizado del problema de comparación de las estimaciones se refería al enfoque estándar. A este último también se refiere, en realidad, lo siguiente

Definición 2. La estimación $\theta_1^* \in K$ se denomina *asintóticamente eficiente* (a.e.) en K si cuando $n \to \infty$, para toda otra estimación θ^* de K y para cada $\theta \in \Theta$,

$$\limsup_{n\to\infty} \frac{\mathbf{M}_\theta(\theta_1^* - \theta)^2}{\mathbf{M}_\theta(\theta^* - \theta)^2} \leqslant 1. \tag{8}$$

Pasemos ahora al enfoque *asintótico* con el cual la definición 2 también está relacionada estrechamente. Aquí, como antes, el problema consiste en la comparación de las funciones $\sigma(\theta)$ que caracterizan la distribución normal límite, pero la cuestión en general se simplifica un poco. Esto se debe, ante todo, a que la comparación se realiza solamente en la clase de estimaciones a.n., que en lo sucesivo la designaremos por K_Φ. Podemos contraer un poco esta clase K_Φ sin empobrecerla considerablemente. Así pues, examinaremos la clase $K_{\Phi,2} \subset K_\Phi$ de las estimaciones a.n. θ^* que poseen la propiedad de que para ellas la convergencia

$$(\theta^* - \theta)\sqrt{n} \Leftarrow \Phi_{0,\sigma^2(\theta)}$$

ocurre junto con los dos primeros momentos:

$$\mathbf{M}_\theta(\theta^* - \theta)\sqrt{n} \to 0, \ \mathbf{M}_\theta(\theta^* - \theta)^2 n \to \sigma^2(\theta). \tag{9}$$

Señalemos que la primera de estas dos relaciones se obtiene fácilmente de

---

[*] Es válida la siguiente afirmación que generaliza, en cierto sentido, el teorema 1. *Si θ_0 es eficiente en k_b y la estimación θ^* es arbitraria en k_b, de modo que $h = D_\theta\theta_0^*/D_\theta\theta^* \leqslant 1$, entonces el coeficiente de correlación $\varrho(\theta_0^*, \theta^*)$ entre las estimaciones θ_0^* y θ^* es igual a*

$$\varrho(\theta_0^*, \theta^*) = \sqrt{h}.$$

El lector puede realizar individualmente la demostración, después de convencerse de que cuando $\varrho(\theta_0^*, \theta^*) \neq \sqrt{h}$ y al elegir correspondiente α, la estimación

$$\theta_1^* = (1 - \alpha)\theta_0^* + \alpha\theta^* \in K_b$$

satisfará la desigualdad $D_\theta\theta_1^* < D_\theta\theta_0^*$ que contradice la eficacia de θ_0^*.

la segunda con ayuda del teorema de continuidad para los momentos (§ 1.5).

La contracción de K_Φ hasta la clase $K_{\Phi,2}$ empobrece poco la primera de estas clases por dos causas. En primer lugar, las estimaciones a.n. en las que (9) no se cumple, prácticamente no existen (hemos señalado que para esto son necesarias, por regla general, construcciones artificiales). En segundo lugar, para $\theta^* \in K$, conforme al lema de Fatou,

$$\liminf_{n\to\infty} \mathbf{M}_\theta n(\theta^* - \theta)^2 \geqslant \sigma^2(\theta)$$

(se trata de las integrales de las funciones no negativas), así que $\mathbf{M}_\theta n(\theta^* - \theta)^2$, para grandes valores de n puede distinguirse de $\sigma^2(\theta)$ únicamente hacia el lado de los valores más grandes. Pero es poco probable que las estimaciones con tales propiedades puedan competir con las estimaciones para las cuales (9) ha sido cumplida.

Ahora bien, cuando se trata del enfoque asintótico, en calidad de clase de estimaciones a.n., en la cual se realiza la comparación, podemos considerar la clase $K_{\Phi,2}$. Esta será más cómoda para nosotros.

Sea K cierta clase de estimaciones, tal que $K \subset K_{\Phi,2}$. Entonces la siguiente definición será equivalente a la definición 2.

Definición 3. La estimación $\theta_1^* \in K$ se llama *asintóticamente eficiente en K, si para cualquier otra estimación* $\theta^* \in$ K

$$\sigma_1^2(\theta) \leqslant \sigma^2(\theta) \tag{10}$$

cuando todos $\theta \in \Theta$, donde $\sigma^2(\theta)$ y $\sigma_1^2(\theta)$ son los coeficientes de dispersión de θ^* y θ_1^*, respectivamente.

La equivalencia de las definiciones se deduce del hecho de que para $\theta^* \in K_{\Phi,2}$

$$\mathbf{M}_\theta(\theta^* - \theta)^2 = \frac{\sigma^2(\theta)}{n}(1 + r_n(\theta)),\ r_n(\theta) \to 0 \text{ cuando } n \to \infty.$$

En este caso la relación (8), que significa que

$$\mathbf{M}_\theta(\theta_1^* - \theta)^2 \leqslant \mathbf{M}_\theta(\theta^* - \theta)^2(1 + r_n'(\theta)),\ r_n'(\theta) \to 0,$$

para cualquier $\theta^* \in$ K es, evidentemente, equivalente a la desigualdad (10). ◁

En el enfoque asintótico, cierta simplificación del problema de comparación (anteriormente recordada) consiste en que aquí comparamos tan sólo las varianzas de las leyes del límite. Aquí desaparece la importancia del desplazamiento $b(\theta)$ de las estimaciones, puesto que en la clase $K_{\Phi,2}$, en virtud de (9) se cumple la relación $b(\theta) = o(1/\sqrt{n})$ que significa "casi la falta de desplazamiento" de las estimaciones o la "despreciabilidad asintóti-

ca" del desplazamiento desde el punto de vista de las relaciones (2).

Análogamente al teorema 1 puede ser obtenido

El teorema 2. *Sea $K \subset K_{\Phi,2}$. Entonces, si θ_1^* y θ_2^* son dos estimaciones a.e. en K, tales que $\frac{1}{2}(\theta_1^* + \theta_2^*) \in K$, éstas coinciden asintóticamente, o sea,*

$$\sqrt{n}(\theta_1^* - \theta_2^*) \underset{P}{\to} 0, \quad \mathbf{M}_\theta[\sqrt{n}(\theta_1^* - \theta_2^*)]^2 \to 0.$$

Demostración. Basta determinar la segunda relación, ya que la primera se deduce de ella. Sea

$$\mathbf{M}_{l,n} = \mathbf{M}_\theta n(\theta_l^* - \theta)^2, \ \Delta_l = \theta_l^* - \theta, \ \theta^* = \frac{\theta_1^* + \theta_2^*}{2}, \ l = 1, 2.$$

Entonces, en virtud de (6) obtenemos

$$\mathbf{M}_\theta n(\theta^* - \theta)^2 + \frac{1}{4}\mathbf{M}_\theta n(\theta_1^* - \theta_2^*)^2 = (M_{1,n} + M_{2,n})/2. \tag{11}$$

Pero $\theta^* \in K$ y, por consiguiente, después de pasar al límite, en la última igualdad obtenemos, en virtud de la eficacia asintótica de θ_l^*,

$$\lim_{n\to\infty} \mathbf{M}_\theta n(\theta_1^* - \theta_2^*)^2 \leqslant 0. \ \triangleleft$$

Las consideraciones expuestas anteriormente contenían sólo una de las vías posibles de separar las estimaciones (en nuestro caso, las estimaciones eficientes) que, siguiendo varios razonamientos naturales, han de preferirse a otras. No obstante, son posibles, desde luego, también otros enfoques (recuérdese que teníamos que comparar los elementos no ordenados, o sea, las funciones $d(\theta)$ o $\sigma(\theta)$). Puesto que, hablando en general, no existen estimaciones con valores mínimos posibles de $d(\theta)$ para cada θ, entonces se pueden comparar, digamos, los valores medios $\int d(t)\,q(t)\,dt$, donde $q(t) \geqslant 0$, $\int q(t)\,dt = 1$, o los valores máximos $\max\limits_{\theta\in\Theta} d(\theta)$. Esto son los métodos de reglamentación de los conjuntos de todas las estimaciones.

Más tarde llamaremos *bayesiano* el primero de estos dos métodos, y *minimax*, el segundo. Las estimaciones óptimas bayesianas y minimax serán examinadas en el § 11, y las estimaciones eficientes, en los párrafos ulteriores.

El problema de elección de las estimaciones será examinado más detalladamente en el capítulo 5.

2. Caso multidimensional. Examinemos ahora el caso cuando θ y θ^* son vectores de R^k. Aquí, el problema de comparación de las estimaciones es más difícil. El hecho es que en el caso multidimensional teníamos que introducir un orden *parcial* ya para comparar las estimaciones cuando θ

ha sido *registrado*. Para comparar las estimaciones *en todo el conjunto* Θ, al igual que en el caso unidimensional, también es necesario introducir un orden parcial, pero ya "en otra dirección" puesto que la comparación se basa en la desviación estándar, que es una función de dos variables: θ y del vector a, sobre el cual se proyecta la desviación $\theta^* - \theta$).

Las mejores estimaciones en "ambas direcciones" constituyen precisamente el objeto de las definiciones siguientes.

Definición 4. La estimación θ_0^* es *eficiente en la clase* K si para cualquier estimación θ^* de K la dispersión estándar de θ^* alrededor de θ para todos $\theta \in \Theta$ no es menor que la dispersión de θ_0^*.

Esta definición es equivalente a la siguiente.

La estimación vectorial θ_0^* del parámetro θ es eficiente en K si para cualquier vector a la estimación $\alpha_0^* = (\theta_0^*, a)$ es la estimación eficiente del parámetro escalar $\alpha = (\theta, a)$ en la clase de estimaciones $\alpha^* = (\theta^*, a)$, $\theta^* \in K$, o sea, para todos $\theta \in \Theta$, $a \in R^k$, $\theta^* \in K$,

$$\mathbf{M}_\theta(\theta_0^* - \theta, a)^2 \leqslant \mathbf{M}_\theta(\theta^* - \theta, a)^2. \tag{12}$$

Como ya hemos visto, esta desigualdad se escribe de un modo equivalente en la forma $d_0^2(\theta) \leqslant d^2(\theta)$ o bien

$$\sum_{i,j} d_{ij}^{(0)}(\theta) a_i a_j \leqslant \sum_{i,j} d_{ij} a_i a_j$$

para todos $\theta \in \Theta$, $a \in R^k$, donde $d^2(\theta) = |d_{ij}(\theta)|$ y $d_0^2(\theta) = |d_{ij}^{(0)}(\theta)|$ son las matrices de segundos momentos $\theta^* - \theta$ y $\theta_0^* - \theta$, respectivamente.

Las estimaciones eficientes en la clase K_0 de las estimaciones no desplazadas se llaman simplemente *eficientes*.

En vista de que la definición (12) de la eficacia se construye a base de la utilización del caso unidimensional, estonces, mediante el teorema 1 no es difícil establecer que la estimación eficiente en la clase K_b de estimaciones, con un desplazamiento $b(\theta) = \mathbf{M}\theta^* - \theta$ registrado, es la única.

La definición de las estimaciones a.e. en el caso multidimensional es análoga a las definiciones 2 y 3.

Definición 5. La estimación vectorial θ_1^* del parámetro θ es *asintóticamente eficiente* en K si para cualquier vector a la estimación (θ_1^*, a) es la estimación a.e. del parámetro escalar $\alpha = (\theta, a)$ en la clase de estimaciones $\alpha^* = (\theta^*, a)$, $\theta^* \in K$.

En otros términos (véase el § 7), la dispersión estándar de la distribución límite $(\theta_1^* - \theta)\sqrt{n}$, para la estimación a.e. es mínima. Esto, a su vez, significa que para cualesquiera $\theta^* \in K$, $a \in R^k$, $\theta \in \Theta$ se cumple $\sigma_1^2(\theta) \leqslant \sigma^2(\theta)$, o bien

$$\sum_{i,j} \sigma_{ij}^{(1)}(\theta) a_i a_j \leqslant \sum_{i,j} \sigma_{ij}(\theta) a_i a_j,$$

donde $\sigma^2(\theta) = \|\sigma_{ij}(\theta)\|$, $\sigma_1^2(\theta) = \|\sigma_{ij}^{(1)}(\theta)\|$ son, respectivamente, las matrices de segundos momentos de las distribuciones límite $(\theta^* - \theta)\sqrt{n}$ y $(\theta_1^* - \theta)\sqrt{n}$.

Del párrafo precedente se puede sacar la conclusión de que el conjunto de estimaciones en el caso multidimensional, para θ registrado, puede ser ordenado si la calidad de la estimación se mide en cantidad (durante el enfoque estándar)

$$v(\theta^*) = \mathbf{M}_\theta(\theta^* - \theta)V(\theta^* - \theta)^T \equiv v(\theta^*, \theta), \tag{13}$$

donde V es la matriz definida no negativamente. La cantidad análoga relacionada con la matriz de segundos momentos de la distribución normal límite, también se puede examinar durante el enfoque asintótico en la clase $K_{\Phi,2}$.

Continuando el avance por este camino, es posible ordenar bien el conjunto de todas las estimaciones incluso en todo el conjunto Θ. A saber, se pueden comparar los valores medios

$$\int v(\theta^*, t)\, q(t)\, dt, \quad q(t) \geqslant 0, \quad \int q(t)\, dt = 1,$$

o los valores máximos $\max\limits_{t \in \Theta} v(\theta^*, t)$ de las cantidades $v(\theta^* \cdot \theta)$ definidas en (13).

Si resulta que la estimación que es la mejor en tal enfoque, continúa siendo la mejor para cualquier matriz V definida no negativamente, esto significará, en virtud del lema 7.1, que esta estimación también será la mejor desde el punto de vista del orden parcial establecido en el § 7 (o sea, la desviación estándar mediada será la mínima en cualquier dirección).

Para construir las estimaciones óptimas en sentido de las definiciones examinadas en este párrafo, necesitaremos los conceptos y las propiedades de las esperanzas matemáticas condicionales y de las estadísticas suficientes.

§ 9. Esperanzas matemáticas condicionales

En este párrafo recordaremos la definición de las esperanzas matemáticas condicionales (e.m.c) y sus propiedades principales. Véase una exposición más completa en el suplemento III, así como en [11], [38[, [30], [61] y [84].

1. Definición de la e.m.c. Sean ξ y η dos variables aleatorias dadas en el espacio probabilístico $(\Omega, \mathfrak{F}, \mathbf{P})$.

La esperanza matemática condicional $\mathbf{M}(\xi/B)$ de la variable aleatoria ξ respecto al suceso B, $\mathbf{P}(B) > 0$, se define por la igualdad

$$\mathbf{M}(\xi/B) = \frac{\mathbf{M}(\xi;\, B)}{\mathbf{P}(B)}, \tag{1}$$

donde $\mathbf{M}(\xi;\ B) = \int_B \xi d\mathbf{P} = \mathbf{M}(\xi I_B)$, $I_B = I_B(\omega)$ es una variable aleatoria igual al indicador del conjunto B.

Admitamos que ξ y η son independientes, $B = \{\eta = x\}$ y $\mathbf{P}(B) > 0$. Entonces, para cualquier función medible $\varphi(x, y)$ conforme a (1),

$$\mathbf{M}[\varphi(\xi, \eta)/\eta = x] = \frac{\mathbf{M}\varphi(\xi, \eta)I_{\{\eta = x\}}}{\mathbf{P}(\eta = x)} = \frac{\mathbf{M}\varphi(\xi, x)I_{\{\eta = x\}}}{\mathbf{P}(\eta = x)} = \mathbf{M}\varphi(\xi, x) \quad (2)$$

La última igualdad es válida, ya que las variables aleatorias $\varphi(\xi, x)$ e $I_{\{\eta = x\}}$ como funciones de ξ y η, respectivamente, son independientes y, por consiguiente,

$$\mathbf{M}\varphi(\xi, x)I_{\{\eta = x\}} = \mathbf{M}\varphi(\xi, x)\mathbf{M}I_{\{\eta = x\}} = \mathbf{M}\varphi(\xi, x)\mathbf{P}(\eta = x).$$

Las relaciones (2) muestran que el concepto de e.m.c. también puede conservar su significado en el caso cuando la probabilidad de la condición es igual a 0: pues de por sí la igualdad

$$\mathbf{M}[\varphi(\xi, \eta)/\eta = x] = \mathbf{M}\varphi(\xi, x)$$

para ξ y η independientes se presenta natural, y con la suposición de $\mathbf{P}(\eta = x) > 0$ no está relacionada de ningún modo.

Supongamos que $\mathfrak{A}$ es la σ-álgebra de $\mathfrak{F}$. Vamos a definir ahora el concepto de e.m.c. de la variable aleatoria ξ con respecto a $\mathfrak{A}$ que designaremos por $\mathbf{M}(\xi/\mathfrak{A})$. Primero daremos la definición del caso "discreto", pero de modo que se generalice fácilmente.

Llamamos "discreto" el caso cuando la σ-álgebra de $\mathfrak{A}$ está formada (generada) no más que por una sucesión numerable de los sucesos disjuntos $A_1, A_2, \ldots$; $\bigcup A_i = \Omega$, $\mathbf{P}(A_i) > 0$. Este hecho se escribe en forma de $\mathfrak{A} = \sigma(A_1, A_2, \ldots)$ y significa que como elementos de $\mathfrak{A}$ sirven todas las uniones posibles de los conjuntos $A_1, A_2, \ldots$

Con ayuda de la variable aleatoria ξ y el sistema de sucesos $(A_1, A_2, \ldots)$ construiremos una nueva variable aleatoria $\hat{\xi} = \hat{\xi}(\omega)$ del modo siguiente:

$$\hat{\xi} = y_k \equiv \mathbf{M}(\xi/A_k) = \frac{\mathbf{M}(\xi;\ A_k)}{\mathbf{P}(A_k)} \text{ cuando } \omega \in A_k,\ k = 1, 2, \ldots$$

Con otras palabras,

$$\hat{\xi} = \sum_k \frac{\mathbf{M}(\xi;\ A_k)}{\mathbf{P}(A_k)} I_{A_k},$$

donde I_A es el indicador del conjunto A.

Definición 1. La variable aleatoria $\hat{\xi}$ se llama *e.m.c. de* ξ *con respecto a la σ-álgebra de* $\mathfrak{A}$ *y se designa por* $\mathbf{M}(\xi/\mathfrak{A})$.

Ahora bien, a distinción de las esperanzas matemáticas ordinarias, la e.m.c. $\mathbf{M}(\xi/\mathfrak{A})$ es una *variable aleatoria*. En nuestro caso esta variable es constante en los conjuntos A_k y equivale, en estos conjuntos, al promedio de ξ en A_k. Si ξ y $\mathfrak{A}$ son independientes (o sea, $\mathbf{P}(\xi \in B; A_k) = \mathbf{P}(\xi \in B)\mathbf{P}(A_k)$), entonces es evidente que $\mathbf{M}(\xi; A_k) = \mathbf{M}\xi\mathbf{P}(A_k)$ y $\hat{\xi} = \mathbf{M}\xi$.

Sin embargo, si $\mathfrak{A} = \mathfrak{F}$, entonces $\mathfrak{F}$ también es "discreta", ξ es constante en los conjuntos A_k y, por lo tanto, $\hat{\xi} = \xi$. Señalemos las dos propiedades principales siguientes de la e.m.c.:

1) $\hat{\xi}$ *es medible con respecto* a $\mathfrak{A}$.

2) *Para cualquier suceso* $A \in \mathfrak{A}$

$$\mathbf{M}(\hat{\xi}; A) = \mathbf{M}(\xi; A).$$

La primera propiedad es evidente. La segunda se deduce del hecho de que todo suceso $A \in \mathfrak{A}$ es representable en la forma $A = \bigcup_k A_{j_k}$ y, por consiguiente,

$$\mathbf{M}(\hat{\xi}; A) = \sum_k \mathbf{M}(\hat{\xi}; A_{j_k}) = \sum_k y_{j_k}\mathbf{P}(A_{j_k}) = \sum_k \mathbf{M}(\xi; A_{j_k}) = \mathbf{M}(\xi; A).$$

Esta propiedad es bastante clara: tras promediar la variable ξ respecto al conjunto A se obtiene el mismo resultado que al promediar la magnitud $\hat{\xi}$ ya promediada respecto a A_{j_k}.

Lema 1. *Las propiedades* 1) y 2) *definen unívocamente la e.m.c. y son equivalentes a la definición 1.*

Demostración. En una dirección la afirmación del lema ya está demostrada. Ahora supongamos que se han cumplido las condiciones 1 y 2. La mensurabilidad de $\hat{\xi}$ con respecto a $\mathfrak{A}$ quiere decir que $\hat{\xi}$ es constante en los conjuntos A_k. Designemos el valor de $\hat{\xi}$ sobre A_k a través de y_k. Como $A_k \in \mathfrak{A}$, de la propiedad 2 se deduce que

$$\mathbf{M}(\hat{\xi}; A_k) = y_k\mathbf{P}(A_k) = \mathbf{M}(\xi; A_k)$$

y, por lo tanto, para $\omega \in A_k$

$$\hat{\xi} = y_k = \frac{\mathbf{M}(\xi; A_k)}{\mathbf{P}(A_k)}. \triangleleft$$

Ahora podemos dar la definición general de la e.m.c.

Definición 2. Supongamos que ξ es una variable aleatoria en el espacio probabilístico $(\Omega, \mathfrak{F}, \mathbf{P})$ y que $\mathfrak{A} \subset \mathfrak{F}$ es la σ-subálgebra de $\mathfrak{F}$. Llámase *esperanza matemática condicional de* ξ *respecto* a $\mathfrak{A}$ la variable aleatoria $\hat{\xi}$ desig-

nada por $\mathbf{M}(\xi/\mathfrak{A})$, la cual posee las dos propiedades siguientes:

1) *$\hat{\xi}$ es medible respecto a $\mathfrak{A}$.*

2) *Para cualquier $A \in \mathfrak{A}$ es válida* $\mathbf{M}(\hat{\xi}; A) = \mathbf{M}(\xi; A)$.

En esta definición la variable aleatoria ξ puede ser tanto escalar como vectorial.

En seguida surgen las preguntas: ¿existe tal variable $\hat{\xi}$? y ¿es única ésta? Hemos visto que en el caso "discreto" la respuesta a estas preguntas es positiva. En el caso general es válido

Teorema 1. *Si* $\mathbf{M}|\xi|$ *es finita, entonces la función* $\hat{\xi} = \mathbf{M}(\xi/\mathfrak{A})$ *siempre existe en la definición 2 y es única con una exactitud de hasta los valores en el conjunto de probabilidad cero.*

Demostración. Primero supongamos que ξ es escalar, $\xi \geqslant 0$. Entonces la función del conjunto

$$\mathbf{Q}(A) = \int_A \xi d\mathbf{P} = \mathbf{M}(\xi; A), \quad A \in \mathfrak{A},$$

será la medida en $(\Omega, \mathfrak{A})$, que es absolutamente continua respecto a $\mathbf{P}$, puesto que $\mathbf{P}(A) = 0$ conduce a $\mathbf{Q}(A) = 0$. Por consiguiente, según el teorema de Radón—Nikodym ([11], Suplemento 3) existe la función $\mathfrak{A}$-medible $\hat{\xi} = \mathbf{M}(\xi/\mathfrak{A})$ única, con una exactitud de hasta los valores en el conjunto de medida cero, tal que

$$\mathbf{Q}(A) = \int_A \hat{\xi} d\mathbf{P}.$$

En el caso general pongamos $\xi = \xi^+ - \xi^-$, $\xi^+ = \text{máx}(0, \xi) \geqslant 0$, $\xi^- = \text{máx}(0, -\xi) \geqslant 0$,

$$\hat{\xi} = \hat{\xi}^+ - \hat{\xi}^-,$$

donde $\hat{\xi}^\pm$ es la e.m.c. para $\xi^\pm$. Esto demuestra la existencia de la e.m.c., ya que $\hat{\xi}$ satisfará las condiciónes 1) y 2) de la definición 2. De aquí también resulta la unicidad, ya que la suposición acerca de la no unicidad de $\hat{\xi}$ significará la no unicidad de $\hat{\xi}^+$ o de $\hat{\xi}^-$. La demostración para ξ vectoriales se reduce al caso unidimensional, ya que las propiedades 1) y 2) pertenecerán a las coordenadas de $\hat{\xi}$ cuya existencia y unicidad ya han sido demostradas. ◁

La esencia de la demostración citada es bastante clara: pues según la condición 2, para cualquier $A \in \mathfrak{A}$ se da $\mathbf{M}(\xi; A) = \int_A \hat{\xi} d\mathbf{P}$, o sea, se dan los valores de las integrales de $\hat{\xi}$ de todos los conjuntos $A \in \mathfrak{A}$. Es evidente que esto debe definir unívocamente la función $\mathfrak{A}$-medible $\hat{\xi}$ con una exactitud de hasta los valores en el conjunto de medida 0.

El sentido de $\mathbf{M}(\xi/\mathfrak{A})$ queda el mismo y, en términos generales, constituye el promedio de ξ en los elementos "indivisibles" de $\mathfrak{A}$.

Si $\mathfrak{A} = \mathfrak{F}$, entonces, evidentemente, $\tilde{\xi} = \xi$ satisface las propiedades 1) y 2) y, por lo tanto, $\mathbf{M}(\xi/\mathfrak{F}) = \xi$.

Definición 3. Supongamos que ξ y η son las variables aleatorias en $(\Omega, \mathfrak{F}, \mathbf{P})$ y que $\mathfrak{A} = \sigma(\eta)$ es la σ-álgebra engendrada por la variable aleatoria η. Entonces $\mathbf{M}(\xi/\mathfrak{A})$ también se llama esperanza matemática condicional de la variable ξ respecto a η.

A veces, para simplificar la exposición, en vez de $\mathbf{M}(\xi/\sigma(\eta))$ escribiremos $\mathbf{M}(\xi/\eta)$, lo cual no conduce a equivocaciones.

Como, por definición, $\mathbf{M}(\xi/\eta)$ es una variable $\sigma(\eta)$-medible aleatoria, esto significa (véase [11], p.65) que existe una función medible $g(x)$ para la cual

$$\mathbf{M}(\xi/\eta) = g(\eta). \tag{3}$$

Por analogía con el caso discreto, la magnitud $g(x)$ aquí puede ser interpretada como el resultado de la mediación de ξ en el conjunto $\{\eta = x\}$. Recordemos que en el caso discreto $g(x) = \mathbf{M}(\xi/\eta = x))$.

Definición 4. Si $\xi = I_C$ es el indicador del conjunto $C \in \mathfrak{F}$, entonces $\mathbf{M}(I_C/\mathfrak{A})$ se denominará *probabilidad condicional* $\mathbf{P}(C/\mathfrak{A})$ del suceso C respecto a $\mathfrak{A}$. Si $\mathfrak{A} = \sigma(\eta)$, entonces hablaremos de la probabilidad condicional $\mathbf{P}(C/\eta)$ del suceso C respecto a η.

Propiedades de la e.m.c.

1) *La e.m.c. posee propiedades de esperanzas matemáticas ordinarias* (véase [11], p.75), con la única diferencia de que las mismas se cumplen casi con seguridad (con probabilidad 1):

1a) $\mathbf{M}(c\xi/\mathfrak{A}) = c\mathbf{M}(\xi/\mathfrak{A})$ *si* c = const,

1b) $\mathbf{M}(\xi_1 + \xi_2/\mathfrak{A}) = \mathbf{M}(\xi_1/\mathfrak{A} + \mathbf{M}(\xi_2)/\mathfrak{A})$,

1c) *si* $\xi_1 \leqslant \xi_2$ c.s., *entonces* $\mathbf{M}(\xi_1/\mathfrak{A}) \leqslant \mathbf{M}(\xi_2/\mathfrak{A})$.

2) *Es válida la desigualdad del tipo de Chébishev*: si ξ es real, $\xi \geqslant 0$, entonces para cualquier $x > 0$,

$$\mathbf{P}(\xi \geqslant x/\mathfrak{A}) \leqslant \frac{\mathbf{M}(\xi/\mathfrak{A})}{x}.$$

Lo mismo que las igualdades del punto 1, tal relación entre las e.m.c. se cumple casi con seguridad. Este mismo acuerdo será válido posteriormente para todas las relaciones entre las e.m.c.

3) *Si las σ-álgebras de $\mathfrak{A}$ y $\sigma(\xi)$ son independientes, entonces* $\mathbf{M}(\xi/\mathfrak{A}) = \mathbf{M}\xi$.

De aquí se deduce, en particular, que si ξ y η son independientes, entonces $\mathbf{M}(\xi/\eta) = \mathbf{M}\xi$. Si la σ-álgebra de $\mathfrak{A}$ es trivial, entonces, evidentemente, también obtenemos $\mathbf{M}(\xi/\mathfrak{A}) = \mathbf{M}\xi$.

4) *Para las e.m.c. son ciertos los teoremas de convergencia, válidos para las esperanzas matemáticas ordinarias,* por ejemplo, el teorema de convergencia monótona: si $\xi_n \uparrow \xi$, $\xi_n \geqslant 0$, entonces $\mathbf{M}(\xi_n/\mathfrak{A}) \uparrow \mathbf{M}(\xi/\mathfrak{A})$ c.s.

5) *Si η es escalar y medible respecto a $\mathfrak{A}$,* $\mathbf{M}|\xi| < \infty$, $\mathbf{M}|\xi_\eta| < \infty$, entonces

$$\mathbf{M}(\eta\xi/\mathfrak{A}) = \eta\mathbf{M}(\xi/\mathfrak{A}).$$

Con otras palabras, las variables aleatorias $\mathfrak{A}$-medibles se comportan, respecto a la operación de e.m.e., como constantes (compararlo con la propiedad 1a).

6) *Para las e.m.c. quedan válidas todas las desigualdades principales para las esperanzas matemáticas ordinarias,* en particular, la desigualdad de Cauchy — Buniakovski

$$\mathbf{M}(|\xi_1\xi_2|/\mathfrak{A}) \leqslant [\mathbf{M}(\xi_1^2/\mathfrak{A})\mathbf{M}(\xi_2^2/\mathfrak{A})]^{1/2}$$

y la desigualdad de Jensen: si $\mathbf{M}|\xi| < \infty$, entonces para cualquier función $g(x)$ convexa hacia abajo,

$$g(\mathbf{M}(\xi/\mathfrak{A})) \leqslant \mathbf{M}(g(\xi)/\mathfrak{A}).$$

7) *Fórmula de la probabilidad completa* (propiedad 2 de la definición 2 cuando $A = \Omega$):

$$\mathbf{M}\xi = \mathbf{M}\mathbf{M}(\xi/\mathfrak{A}).$$

8) *Promediación sucesiva* (generalización de la propiedad 7)): si $\mathfrak{A} \subset \mathfrak{A}_1 \subset \mathfrak{F}$, entonces

$$\mathbf{M}(\xi/\mathfrak{A}) = \mathbf{M}(\mathbf{M}(\xi/\mathfrak{A}_1)/\mathfrak{A}).$$

En el Suplemento III se puede hallar la demostración de estas propiedades.

Es evidente que las propiedades 1), 3), — 5), 7) y 8) son válidas tanto para las variables aleatorias ξ escalares como para las vectoriales. Destacaremos especialmente la siguiente propiedad de las e.m.c.

9) Es sabido que la función $\varphi(\alpha) = \mathbf{M}(\xi - a)^2$ alcanza su valor mínimo cuando $a = \mathbf{M}\xi$ (véase, por ejemplo, [11]). Esa misma propiedad también es válida para la e.m.c.: *cuando $a(\omega) = \mathbf{M}(\xi/\mathfrak{A})$ se alcanza el valor mínimo $\mathbf{M}(\xi - a(\omega))^2$ entre todas las funciones $a(\omega)$ $\mathfrak{A}$-medibles.*

En efecto, $\mathbf{M}(\xi - a(\omega))^2 = \mathbf{M}\mathbf{M}((\xi - a(\omega))^2/\mathfrak{A}$, pero $a(\omega)$ se comporta como constante respecto a la operación $\mathbf{M}(\cdot/\mathfrak{A})$ (véase la propiedad 5)), así que

$$\mathbf{M}((\xi - a(\omega))^2/\mathfrak{A}) = \mathbf{M}((\xi - \mathbf{M}(\xi/\mathfrak{A}))^2/\mathfrak{A}) + \mathbf{M}((\mathbf{M}(\xi/\mathfrak{A}) - a(\omega))^2/\mathfrak{A})$$

y el valor mínimo de esta expresión se alcanza cuando $a(\omega) = \mathbf{M}(\xi/\mathfrak{A})$. Esta propiedad puede considerarse como definición de la e.m.c. equivalente a

la definición 2. Debido a ella, $\mathbf{M}(\xi/\mathfrak{A})$ puede interpretarse como la "proyección" de ξ sobre $\mathfrak{A}$.

La propiedad 9) admite la siguiente generalización para el caso multidimensional, cuando $\xi = (\xi_1, \ldots, \xi_s)$ es un vector aleatorio en R^s.

9A) *Sea* $V = \|v_{ij}\|$ *una matriz arbitraria, definida no negativamente y de dimensión* $s \times s$, $a \in R^s$,

$$\zeta(a) = (\xi - a)V(\xi - a)^T$$

(en particular, para $V = E$ obtenemos $\zeta(a) = |\xi - a|^2$). *Entonces, en la función* $a(\omega) = \mathbf{M}(\xi/\mathfrak{A})$ *se alcanza el valor mínimo* $\min_{a \in A} \mathbf{M}\zeta(a)$ *para la clase* A *de todas las funciones* $\mathfrak{A}$*-medibles.*

La demostración de este hecho transcurre igual que en el caso unidimensional. Designemos $\alpha = \mathbf{M}(\xi/\mathfrak{A})$. Entonces $\mathbf{M}\zeta(a) = \mathbf{M}\mathbf{M}(\zeta(a)/\mathfrak{A})$,

$$\begin{aligned}\mathbf{M}(\zeta(a)/\mathfrak{A}) = \mathbf{M}((\xi - a)V(\xi - a)^T/\mathfrak{A} = \mathbf{M}((\xi - \alpha)V(\xi - \alpha)^T/\mathfrak{A}) + \\ + \mathbf{M}((\alpha - a)V(\xi - \alpha)^T/\mathfrak{A}) + \mathbf{M}((\xi - \alpha)V(\alpha - a)^T/\mathfrak{A}) + \\ + \mathbf{M}((\alpha - a)V(\alpha - a)^T/\mathfrak{A}).\end{aligned} \quad (4)$$

Como $\alpha - a$ es el vector $\mathfrak{A}$-medible, entonces, según la propiedad 5),

$$\mathbf{M}((\alpha - a)V(\xi - \alpha)^T/\mathfrak{A}) = (\alpha - a)V\mathbf{M}((\xi - \alpha)^T/\mathfrak{A}) = 0,$$
$$\mathbf{M}((\xi - \alpha)V(\alpha - a)^T/\mathfrak{A}) = [\mathbf{M}((\xi - \alpha)/\mathfrak{A})]V(\alpha - a)^T = 0.$$

En vista de que el último sumando en (4) no es negativo y equivale a cero cuando $a = \alpha$, la afirmación queda demostrada. ◅

§ 10. Distribuciones condicionales

A la par con las e.m.c., las distribuciones condicionales se pueden examinar respecto a las σ-subálgebras y respecto a las variables aleatorias. En este párrafo estudiaremos solamente las distribuciones condicionales respecto a las variables aleatorias.

Sean ξ y η dos variables aleatorias en $(\Omega, \mathfrak{F}, \mathbf{P})$ con valores en R^s y R^k, respectivamente, y sea $\mathfrak{B}^s$ la σ-álgebra de los conjuntos de Borel de R^s.

Definición 1. *La función* $\mathbf{P}(B/y)$ *de dos variables* $y \in R^k$, $B \in \mathfrak{B}^s$ *se llama distribución condicional de* ξ, *a condición de que* $\eta = y$, *si*

1) Para cada B $\mathbf{P}(B/\eta)$ es la probabilidad condicional $\mathbf{P}(\xi \in B/\eta)$ del suceso $\{\xi \in B\}$ respecto a η, o sea, $\mathbf{P}(B/y)$ es una función de Borel de y, tal que para cualquier $A \in \mathfrak{B}^k$,

$$\mathbf{M}(\mathbf{P}(B/\eta); \eta \in A) = \int_A \mathbf{P}(B/y)\mathbf{P}(\eta \in dy) = \mathbf{P}(\xi \in B, \eta \in A).$$

2) Para cada y, $\mathbf{P}(B/y)$ es la distribución de las probabilidades sobre B.

A veces escribiremos la función $\mathbf{P}(B/y)$ de una "forma más descodificada":

$$\mathbf{P}(B/y) = \mathbf{P}(\xi \in B/\eta = y).$$

Sabemos que para cada $B \in \mathfrak{B}^s$ existe una función de Borel $g_B(y)$ tal que $g_B(\eta) = \mathbf{P}(\xi \in B/\eta)$. Ahora bien, poniendo $\mathbf{P}(B/y) = g_B(y)$, satisfaremos la condición 1) de la definición. Sin embargo, en este caso la condición 2) no se deduce de ningún modo de las propiedades de la e.m.c. y de ninguna manera se ve obligada a ser cumplida: pues la probabilidad condicional $\mathbf{P}(\xi \in B/\eta)$ está definida para cada B, con una exactitud de hasta los valores en el conjunto N_B de medida cero (ya que existen muchas variantes de e.m.c.) y este conjunto puede ser propio para cada B. Por eso, si la unión $N = \bigcup_{B \in \mathfrak{B}^s} N_B$ no tiene probabilidad nula, puede resultar que, por ejemplo, las igualdades

$$\mathbf{P}(\xi \in B_1 \cup B_2/\eta) = \mathbf{P}(\xi \in B_1/\eta) + \mathbf{P}(\xi \in B_2/\eta)$$

(aditividad de la probabilidad) a la vez para todos B_1, B_2 disjuntos de $\mathfrak{B}^s$ no se cumplen ni siquiera para un solo ω de N, o sea, en el ω-conjunto de N de una probabilidad positiva, la función $g_B(y)$ no será una distribución como la función B.

No obstante, en nuestro caso, *cuando ξ es una variable aleatoria con valores en R^s y con σ-álgebra de los conjuntos de Borel $\mathfrak{B}^s$, $g_B(\eta) = \mathbf{P}(\xi \in B/\eta)$, siempre se puede elegir de tal modo que $g_B(y)$ sea una distribución condicional* (véase [38], [30]).

Como era de esperar, las distribuciones condicionales poseen la propiedad natural consistente en que las e.m.c. se expresan en forma de integrales según las distribuciones condicionales.

Teorema 1. *Para toda función medible $g(x)$ que aplica R^s en R, tal que $\mathbf{M}|g(\xi)| < \infty$, es válida la igualdad*

$$\mathbf{M}(g(\xi)/\eta) = \int g(x)\mathbf{P}(dx/\eta). \tag{1}$$

Demostración. Es suficiente examinar el caso cuando $g(x) \geqslant 0$. Si $g(x) = I_A(x)$ es el indicador del conjunto A, entonces la fórmula (1) es evidentemente cierta, o sea, es cierta para cualquier función simple $g_n(x)$ (es decir, para una función que adopte un número finito de valores). Nos queda tomar la sucesión $g_n \uparrow g$ y utilizar la monotonía de ambos miembros en (1) y la propiedad 4) del § 9. ◁

En los problemas reales, para calcular las distribuciones condicionales, a menudo es posible valerse de la siguiente regla simple, que, para eviden-

ciar, podemos escribirla de la forma siguiente:

$$\mathbf{P}(\xi \in B/\eta = y) = \frac{\mathbf{P}(\xi \in B, \eta \in dy)}{\mathbf{P}(\eta \in dy)}. \tag{2}$$

Por supuesto que ambas condiciones de la definición 1 serán satisfechas formalmente.

Si ξ y η tienen densidad de distribución, dicha igualdad adquirirá un sentido exacto.

Definición 2. Supongamos que la distribución condicional $\mathbf{P}(B/y)$, para cada y es absolutamente continua respecto a cierta medida μ en R^s:

$$\mathbf{P}(\xi \in B/\eta = y) = \int_B f(x/y)\mu(dx).$$

Entonces la densidad $f(x/y)$ se denomina *densidad condicional de* ξ (*respecto a la medida* μ), *a condición de que* $\eta = y$.

En otros términos, la función $f(x/y)$ medible conforme al par de variables x, y es la densidad condicional de ξ a condición de que $\eta = y$, si

1) Para cualesquiera conjuntos de Borel, $A \subset R^k$, $B \subset R^s$

$$\int_{y \in A} \int_{x \in B} f(x/y)\mu(dx)\mathbf{P}(\eta \in dy) = \mathbf{P}(\xi \in B, \eta \in A), \tag{3}$$

2) Para cada y la función $f(x/y)$ es la densidad de distribución de las probabilidades.

Del teorema 1 se deduce que si existe la densidad condicional, entonces

$$\mathbf{M}(g(\xi)/\eta) = \int g(x)f(x/\eta)\mu(dx).$$

Si suponemos adicionalmente que la distribución de η tiene una densidad $q(y)$ respecto a cierta medida λ en R^k, entonces (3) se puede escribir de la forma siguiente:

$$\int_{y \in A} \int_{x \in B} f(x/y)q(y)\mu(dx)\lambda(dy) = \mathbf{P}(\xi \in B, \eta \in A), \tag{4}$$

Examinemos ahora el producto directo de los espacios R^s y R^k y, a base de él, el producto directo de las medidas $\mu \times \lambda$ (si $C = B \times A$, $B \subset R^s$, $A \subset R^k$, entonces $\mu \times \lambda(C) = \mu(B)\lambda(A)$). En este espacio la relación (4) significa, evidentemente, que la distribución compatible de ξ y η en $R^s \times R^k$ tiene una densidad respecto a $\mu \times \lambda$, igual a

$$f(x, y) = f(x/y)q(y).$$

Pero también es válida la afirmación inversa.

Teorema 2. *Si la distribución compatible de ξ y η en $R^s \times R^k$ tiene una densidad $f(x, y)$ respecto a $\mu \times \lambda$, entonces la función*

$$f(x/y) = \frac{f(x, y)}{q(y)}, \quad \text{donde } q(y) = \int f(x, y)\mu(dx)$$

es la densidad condicional de ξ, a condición de que $\eta = y$, y la función $q(y)$ es la densidad de η respecto a la medida λ.

Demostración. La afirmación del teorema respecto a $q(y)$ es evidente, ya que $\int_A q(y)\lambda(dy) = \mathbf{P}(\eta \in A)$. Queda señalar que $f(x/y) = f(x, y)/q(y)$ satisface todas las condiciones en la definición 2 de la densidad condicional (la igualdad (4) equivalente a 3 está cumplida de un modo evidente). ◁

Observación 1. Las variables aleatorias ξ y η en el teorema 2 se pueden cambiar de lugar. Entonces obtendremos que, a la par con $f(x/y)$, existe la densidad condicional

$$q(y/x) = \frac{f(x, y)}{f(x)}, \; f(x) = \int f(x, y)\lambda(dy)$$

de la variable aleatoria η, a condición de que $\xi = x$. Este simple corolario del teorema 2 desempeñará un papel muy importante en la exposición posterior. Con arreglo a los problemas de la estadística, este corolario nos permitirá obtener, en el párrafo siguiente, la fórmula de Bayes que luego se utilizará con frecuencia a lo largo de todo este curso.

Ejemplo 1. Sea Φ_{α,σ^2} la distribución normal bidimensional de las variables ξ_1 y ξ_2, donde $\alpha = (\alpha_1, \alpha_2)$, $\alpha_i = \mathbf{M}\xi_i$, $\sigma^2 = \|\sigma_{ij}\|$, $\sigma_{ij} = \mathbf{M}(\xi_i - \alpha_i)(\xi_j - \alpha_j)$, $i, j = 1, 2$. El determinante de la matriz de segundos momentos es igual a

$$|\sigma^2| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1 - \varrho^2),$$

donde ϱ es el coeficiente de correlación entre ξ_1 y ξ_2. Ahora bien, si $|\varrho| \neq 1$, la matriz de segundos momentos no está degenerada y para ella existe la matriz inversa

$$A = (\sigma^2)^{-1} = \frac{1}{|\sigma^2|}\begin{Vmatrix} \sigma^{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{Vmatrix} = \frac{1}{1-\varrho^2}\begin{Vmatrix} \dfrac{1}{\sigma_{11}} & -\dfrac{\varrho}{\sqrt{\sigma_{11}\sigma_{22}}} \\ -\dfrac{\varrho}{\sqrt{\sigma_{11}\sigma_{22}}} & \dfrac{1}{\sigma_{22}} \end{Vmatrix}.$$

Por lo tanto, la densidad compatible de ξ_1 y ξ_2 (respecto a la medida de Lebesgue) es igual a (véase el § 2)

$$f(x, y) = \frac{1}{2\pi\sigma_{11}\sigma_{22}\sqrt{1-\varrho^2}} \times$$

$$\times \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\frac{(x-\alpha_1)^2}{\sigma_{11}} - \frac{2\varrho(x-\alpha_1)(y-\alpha_2)}{\sqrt{\sigma_{11}\sigma_{22}}} + \frac{(y-\alpha_2)^2}{\sigma_{22}}\right]\right\}.$$

Las densidades unidimensionales de ξ_1 y ξ_2 son, respectivamente, iguales a

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_{11}}} e^{-\frac{(x-\alpha_1)^2}{2\sigma_{11}}}, \quad q(y) = \frac{1}{\sqrt{2\pi\sigma_{22}}} e^{-\frac{(y-\alpha_2)^2}{2\sigma_{22}}}.$$

Por eso la densidad condicional de ξ_1, a condición de que $\xi_2 = y$, es igual a

$$f(x/y) = \frac{f(x, y)}{q(y)} =$$

$$= \frac{1}{\sqrt{2\pi\sigma_{11}(1-\varrho^2)}} \exp\left\{-\frac{1}{2\sigma_{11}(1-\varrho^2)}\left(x-\alpha_1-\varrho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\,(y-\alpha_2)\right)^2\right\};$$

ésta es la densidad de la distribución normal, con un valor medio $\alpha_1 + \varrho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\,(y-\alpha_2)$ y la varianza $\sigma_{11}(1-\varrho^2)$. De aquí se deduce, en particular, que la e.m.c. de ξ_1 con respecto a ξ_2 es igual a

$$\mathbf{M}(\xi_1/\xi_2) = \alpha_1 + \varrho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\,(\xi_1 - \alpha_2).$$

La recta $x = \alpha_1 + \varrho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\,(y-\alpha_2)$ se llama línea de regresión de ξ_1 sobre ξ_2. La misma proporciona la mejor aproximación estándar de la variable ξ_1 para una $\xi_2 = y$ dada.

Ejemplo 2. Examinemos el problema consistente en calcular la densidad de la variable aleatoria $\xi = \varphi(\zeta, \eta)$, donde ζ y η son independientes. De la fórmula (3), cuando $A = R^k$, resulta que la densidad $f(x)$ de la distribución de ξ se expresa, mediante la densidad condicional $f(x/y)$, por la igualdad

$$f(x) = \int f(x/y)\mathbf{P}(\eta \in dy). \qquad (5)$$

Con arreglo al problema sujeto a examen, por $f(x/y)$ es necesario entender la densidad de la variable aleatoria $\varphi(\zeta, y)$, puesto que $\mathbf{P}(\xi \in B/\eta = y) = \mathbf{P}(\varphi(\zeta, y) \in B)$.

La fórmula (5) suele ser muy útil al calcular las distribuciones de diferentes estadísticas. Por ejemplo, en el punto 6 del § 2 podríamos escribir directamente la fórmula (2.7) para la densidad de la distribución de Fisher sin deducirla de la forma de la función de distribución.

§ 11. Enfoques bayesiano y minimax de la estimación de los parámetros

La esencia del enfoque bayesiano consiste en que el parámetro desconocido θ se examina como *variable aleatoria* con cierta densidad (conocida o desconocida) de distribución $q(t)$, $t \in \Theta$, respecto a la medida λ, la cual, al igual que la medida μ en la condición (A_μ), será lo más a menudo la medida de Lebesgue o la medida de cálculo. La densidad $q(t)$ se llama densidad *a priori*, o sea, dada *antes* del experimento. El enfoque bayesiano supone que el parámetro desconocido θ se ha escogido aleatoriamente de la distribución de densidad $q(t)$.

Supongamos a continuación, que $f_t(x)$, $t \in \Theta$, $x \in \mathscr{X}^n$ es la función de verosimilitud introducida por nosotros en el § 6. Como ya hemos señalado, $f_t(x)$ es, para cada t, la densidad de distribución en $\mathscr{X}^n$. Por eso la función

$$f(x,\ t) = f_t(x)q(t)$$

es la densidad de cierta distribución en $\mathscr{X}^n \times \Theta$ respecto a la medida $\mu^n \times \lambda$ que puede interpretarse como la *densidad de distribución compatible de* X y θ. Con tal enfoque, en virtud del teorema 10.2, la función $f_t(x)$, $x \in \mathscr{X}^n$ *es la densidad condicional de* X a condición de que $\theta = t$:

$$f_t(x) = f(x/t), \quad \mathbf{M}_\theta g(X) = \mathbf{M}(g(X)/\theta).$$

En estos planteamientos, el aspecto formal del asunto exige que $f_t(x)$ sea una función medible en t y x. En lo sucesivo, por doquier donde esto sea necesario, supondremos que dicha propiedad tiene lugar.

Posteriormente, el parámetro, como variable aleatoria, siempre será designado por θ, mientras que para los valores registrados del parámetro utilizaremos las designaciones t, u, etc., así que

$$\mathbf{M}_t g(X) = \mathbf{M}(g(X)/\theta = t).$$

A la par con $f(x/t)$ podemos escribir la *densidad condicional* $q(t/x)$ *de la variable* θ *a condición de que* $X = x$:

$$q(t/x) = \frac{f_t(x)q(t)}{f(x)}, \quad f(x) = \int f_t(x)q(t)\lambda(dt). \tag{1}$$

Esta densidad define la llamada distribución *a posteriori* (o sea, *después del experimento*) de θ, que designaremos por $\mathbf{Q}_x$. La igualdad (1) se denomina *fórmula de Bayes* para la densidad de la distribución a posteriori. En lo sucesivo esta fórmula desempeñará un papel muy importante.

Con arreglo al caso bayesiano, la propiedad 9 de la e.m.c. significa lo siguiente: entre todas las funciones $\theta^* = \varphi(X)$ la mejor estimación para θ

(desde el punto de vista de minimización de $\mathbf{M}(\theta - \varphi(X))^2$) es la función

$$\theta_Q^* = \mathbf{M}(\theta/X) = \int tq(t/X)\lambda(dt) = \int t\mathbf{Q}_X(dt). \tag{2}$$

Definición 1. La estimación θ_Q^* definida por las fórmulas (2) y (1) se llama *bayesiana, correspondiente a la distribución a priori* $\mathbf{Q}$ *de densidad* $q(t)$.

Señalemos una vez más, que para la estimación bayesiana, la desviación estándar incondicional

$$\mathbf{M}(\theta^* - \theta)^2 = \mathbf{MM}((\theta^* - \theta)^2/\theta = \mathbf{MM}_\theta(\theta^* - \theta)^2 = \\ = \int \mathbf{M}_t(\theta^* - t)^2 q(t)\lambda(dt) \tag{3}$$

adopta el valor mínimo posible. La relación (3) muestra que la estimación bayesiana minimiza el valor medio (con una función ponderal dada $q(t)\lambda(dt)$) de la magnitud $\mathbf{M}_t(\theta^* - t)^2$.

Con otras palabras, si θ se escoge al azar, con densidad $q(t)$, *entonces la estimación bayesiana es la mejor desde el punto de vista del enfoque estándar.* La desviación estándar (3) de la estimación bayesiana puede representarse en la forma (véase (1)):

$$\mathbf{M}(\theta_Q^* - \theta)^2 = \int \mathbf{M}_t(\theta_Q^* - t)^2 q(t)\lambda(dt) = \\ = \iint (t - \theta_Q^*)^2 f_t(x) q(t)\lambda(dt)\mu^n(dx) = \int \sigma_{Q_X}^2 f(x)\mu^n(dx) = \mathbf{M}\sigma_{Q_X}^2,$$

donde $\sigma_{Q_X}^2$ es la varianza de la distribución *a posteriori* $\mathbf{Q}_X$:

$$\sigma_{Q_X}^2 = \int (t - \theta_Q^*)^2 q(t/X)\lambda(dt) = \int (t - \mathbf{M}(\theta/X))^2 \mathbf{Q}_X(dt). \tag{4}$$

El otro enfoque de la comparación de las estimaciones, que ya hemos señalado en el § 8, se basa en la comparación $\sup\limits_{t\in\Gamma} \mathbf{M}_t(\theta^* - t)^2$, donde $\Gamma \subset \Theta$ es un subconjunto dado de Θ(Γ coincide con Θ o es igual a aquella de sus partes respecto a la cual se ha logrado determinar que $\theta \in \Gamma$).

Definición 2. La estimación $\bar\theta^*$ se denomina *minimax* si para cualquier otra estimación θ^*

$$\sup_{t\in\Gamma} \mathbf{M}_t(\bar\theta^* - t)^2 \leqslant \sup_{t\in\Gamma} \mathbf{M}_t(\theta^* - t)^2.$$

Con otras palabras, para la estimación minimax se alcanza

$$\inf_{\theta^*} \sup_{t\in\Gamma} \mathbf{M}_t(\theta^* - t)^2 = \sup_{t\in\Gamma} \mathbf{M}_t(\bar\theta^* - t)^2. \tag{5}$$

Establezcamos ciertas relaciones útiles entre las estimaciones bayesianas y minimax.

Teorema 1. *Designemos por* θ_Q^* *la estimación bayesiana para la distribución a priori* $\mathbf{Q}$ *de densidad* q. *Si existe la estimación* θ_1^* *y la distribución*

Q *tales que para todos t*

$$\mathbf{M}_t(\theta_1^* - t)^2 \leqslant \int \mathbf{M}_u(\theta_Q^* - u)^2 q(u)\lambda(du), \tag{6}$$

la estimación θ_1^* es de tipo minimax.

Demostración. Sea θ^* cualquier otra estimación. Entonces $\sup_t \mathbf{M}_t(\theta^* - t)^2 \geqslant \int \mathbf{M}_t(\theta^* - t)^2 q(t)\lambda(dt) \geqslant \int \mathbf{M}_t(\theta_Q^* - t)^2 q(t)\lambda(dt) \geqslant \mathbf{M}_t(\theta_1^* - t)^2$. ◁

Nótese que casi para todos t pertenecientes al portador $N_Q = \{t: q(t) > 0\}$ de la distribución **Q**, en la desigualdad (6) debe cumplirse indispensablemente la igualdad, ya que de lo contrario obtendríamos

$$\int \mathbf{M}_t(\theta_1^* - t)^2 q(t)\lambda(dt) < \int \mathbf{M}_t(\theta_Q^* - t)^2 q(t)\lambda(dt)$$

lo cual contradice la definición de la estimación bayesiana.

Esta observación nos permite enunciar el siguiente criterio del carácter minimax de la estimación, equivalente al teorema 1.

Teorema 2. *Si la estimación θ^**

1) *es bayesiana para cierta distribución* **Q**,

2) $\mathbf{M}_t(\theta^* - t)^2 = c = \text{const}$ *para* $t \in N_Q$,

3) $\mathbf{M}_t(\theta^* - t)^2 \leqslant c$ *para los demás t, entonces θ^* es una estimación minimax.*

Si $\theta^* = \theta_Q^* = \bar{\theta}^*$ satisface este criterio, es evidente que

$$\sup_t \mathbf{M}_t(\bar{\theta}^* - t)^2 = \int \mathbf{M}_t(\theta^* - t)^2 q(t)\lambda(dt). \tag{7}$$

Ahora bien, la estimación minimax es una estimación bayesiana que "iguala" los errores $\mathbf{M}_t(\bar{\theta}^* - t)^2$ para diferentes t. Esto quiere decir que la distribución a priori $\overline{\mathbf{Q}}$, correspondiente a dicha estimación, obliga a ser igualmente atentos a todos los valores posibles de θ sin orientarse, como lo hacen las estimaciones bayesianas θ_Q^* correspondientes a otras distribuciones a priori $\mathbf{Q} \neq \overline{\mathbf{Q}}$, hacia ciertos valores destacados (más probables) de θ. En vista de que en el último caso utilizamos una información complementaria acerca de θ, es natural que para $\mathbf{Q} \neq \overline{\mathbf{Q}}$ las estimaciones θ_Q^* posean desviaciones estándar incondicionales de menores valores:

$$\int \mathbf{M}_t(\theta_Q^* - t)^2\, \mathbf{Q}(dt) \leqslant \int \mathbf{M}_t(\theta_{\bar{Q}}^* - t)^2 \overline{\mathbf{Q}}(dt).$$

Por eso la distribución $\overline{\mathbf{Q}}$ en el teorema 2, la cual corresponde a la estimación minimax $\bar{\theta}^*$, a menudo se llama distribución *pésima*.

En vista de que tal distribución pésima $\overline{\mathbf{Q}}$ no siempre existe (eso suele suceder en los casos cuando Θ es un conjunto ilimitado), se puede proponer el siguiente criterio modificado para determinar la estimación minimax.

Teorema 3. *Si existe la estimación θ_1^* y la sucesión de distribuciones*

$Q^{(k)}$ *con densidades* $q^{(k)}$ *tales que para todos* t

$$M_t(\theta_1^* - t)^2 \leqslant \lim_{k\to\infty} \sup \int M_t(\theta_Q^*(k) - t)^2 q^{(k)}(t)\lambda(dt),$$

entonces la estimación θ_1^* *es minimax.*

La demostración de este teorema es igualmente simple. Para toda estimación θ^* es válida

$$\sup_t M_t(\theta^* - t)^2 \geqslant \int M_t(\theta^* - t)^2 q^{(k)}(t)\lambda(dt) \geqslant \int M_t(\theta^*_{Q^{(k)}} - t)^2 q^{(k)}(t)\lambda(dt).$$

De aquí se deduce que

$$\sup_t M_t(\theta^* - t)^2 \geqslant \lim_{k\to\infty} \sup \int M_t(\theta^*_{Q^{(k)}} - t)^2 q^{(k)}(t)\lambda(dt) \geqslant M_t(\theta_1^* - t)^2. \quad \triangleleft$$

Ejemplo 1. Sea $X \in \Phi_{\alpha,1}$. Determinemos qué representa la estimación bayesiana $\alpha^*_{Q^{(k)}}$ del parámetro α con una distribución normal a priori $Q^{(k)} = \Phi_{(0,k)}$. En este caso debemos poner $\lambda(dt) = dt$,

$$q^{(k)}(t) = \frac{1}{\sqrt{2\pi k}} e^{-\frac{t^2}{2k}}.$$

La distribución a posteriori $Q_X^{(k)}$ tendrá una densidad $q^{(k)}(t/X)$ proporcional (como función de t) a $q^{(k)}(t)f_t(X)$ o bien, que es lo mismo, proporcional a

$$\exp\left\{-\frac{t^2}{2k} - \frac{1}{2}\sum(x_i - t)^2\right\}.$$

De la igualdad

$$-\frac{t^2}{2}\left(\frac{1}{k} + n\right) + \bar{x}nt = -\frac{1}{2}\left(\frac{1}{k} + n\right)\left(t - \frac{\bar{x}n}{\frac{1}{k} + n}\right)^2 + \frac{(\bar{x}n)^2}{2\left(\frac{1}{k} + n\right)}$$

se deduce que

$$Q_X^{(k)} = \Phi_{\frac{\bar{x}nk}{1+nk},\ \frac{k}{1+nk}}.$$

Como la estimación bayesiana $\alpha^*_{Q^{(k)}}$ del parámetro α es igual a la esperanza matemática de la distribución a posteriori, de aquí obtenemos

$$\alpha_{Q^{(k)}} = \frac{\bar{x}nk}{1 + nk} = \frac{\bar{x}}{1 + \frac{1}{nk}}.$$

La varianza de la distribución a posteriori $\sigma^2_{Q_X^{(k)}} = \dfrac{k}{1 + nk}$ no depende

de X. Por consiguiente, en virtud de (4) el error estándar de la estimación bayesiana es igual a

$$\frac{k}{1 + nk} \to \frac{1}{n}$$

cuando $k \to \infty$. Por eso para la estimación $\alpha^* = \bar{x}$ tenemos

$$\mathbf{M}_t(\bar{x} - t)^2 = \frac{1}{n} = \lim_{k \to \infty} \int \mathbf{M}_t(\alpha^*_{Q^{(k)}} - t)^2 q^{(k)}(t)dt$$

y, por lo tanto, según el teorema 3, la estimación $\alpha^* = \bar{x}$ es minimax. La distribución "pésima" sería aquí la distribución uniforme en toda la recta (distribución "límite" para $\Phi_{0,k}$), si tal distribución existiera *).

En el ejemplo siguiente, el conjunto Θ es compacto y existe la distribución "pésima".

Ejemplo 2. Supongamos que $X \in B_p$, o sea, que x_j, $j = 1, \ldots, n$ adoptan los valores 1 y 0, respectivamente, con probabilidades p y $1 - p$, $p \in \Theta = [0, 1]$. Como sabemos, en este caso para la estimación $p^* = \bar{x}$ es válida

$$\mathbf{M}_p(\bar{x} - p)^2 = p(1 - p)/n,$$

así que el criterio del teorema 2 no se ha cumplido. Examinemos la estimación

$$p^* = \frac{\bar{x} + \dfrac{1}{2\sqrt{n}}}{1 + \dfrac{1}{\sqrt{n}}}. \tag{8}$$

Para ella el error

$$\mathbf{M}_p(p^* - p)^2 = \left(1 + \frac{1}{\sqrt{n}}\right)^{-2} \mathbf{M}_p\left(\bar{x} - p + \frac{1}{2\sqrt{n}} - \frac{p}{\sqrt{n}}\right)^2 =$$

$$= \frac{n}{(1 + \sqrt{n})^2}\left(\frac{p(1 - p)}{n} + \frac{(1 - 2p)^2}{4n}\right) = \frac{1}{4(1 + \sqrt{n})^2}$$

no depende de p. Si ahora nos convencemos de que la estimación (8) es bayesiana, determinaremos de este modo su carácter minimax. Examinemos la distribución a priori $\mathbf{Q} = \mathbf{B}_{N+1,N+1}$, donde $\mathbf{B}_{\lambda_1,\lambda_2}$ es la distribución

---

*) Es interesante anotar que la estimación $\alpha^* = \bar{x}$ deja de poseer la propiedad mencionada, si x es una muestra de una distribución normal multidimensional cuya dimensión constituye más de dos ($x_j \in R^k$, $\alpha \in R^k$, $k \geq 3$). Esto se expone más detalladamente en [48].

beta de densidad (véase el punto 8 del § 2)

$$\frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} t^{\lambda_1 - 1}(1 - t)^{\lambda_2 - 1}.$$

Entonces, como

$$f_t(X) = t^{\bar{x}n}(1 - t)^{n(1 - \bar{x})}, \quad q(t) = \frac{\Gamma(2N + 2)}{\Gamma^2(N + 1)} t^N(1 - t)^N,$$

la distribución a posteriori tendrá una densidad $q(t/X)$ que, como función de t, será proporcional a $f_t(X)q(t)$ o bien, que es lo mismo, será proporcional a

$$t^{N + \bar{x}n}(1 - t)^{N + (1 - \bar{x})n}.$$

Esto significa que la distribución a posteriori coincide con $B_{N + \bar{x}n + 1, N + n(1 - \bar{x}) + 1}$. En vista de que el valor medio de la distribución B_{λ_1, λ_2} es igual a $\lambda_1/\lambda_1 + \lambda_2)$ (véase el punto del § 2), la estimación bayesiana p_Q, correspondiente a Q, será igual a

$$p_Q^* = \frac{N + \bar{x}n + 1}{2N + n + 2} = \frac{\bar{x} + (N + 1)/n}{1 + 2(N + 1)/n}.$$

Cuando $N + 1 = \sqrt{n}/2$, está estimación coincidirá con la estimación p^* definida en (8) y, en virtud del teorema 2, será minimax. La distribución Q será la peor (pésima), ya que se concentra o medida que crece n alrededor del "peor" valor del parámetro $p = 1/2$ con el que la varianza de la estimación $\bar{x}$, igual a $p(1 - p)/n = 1/(4n)$, será máxima. La propia estimación $\bar{x}$ no es minimax, ya que

$$\sup_p \frac{p(1 - p)}{n} = \frac{1}{4n} > \frac{1}{4(1 + \sqrt{n})^2}.$$

Al mismo tiempo es natural que para todos los valores de p que están fuera del entorno estrecho del punto $p = 1/2$, la estimación $\bar{x}$ será, sin embargo, mejor que p_Q^*, y esto tendrá lugar para todos los valores p para los cuales

$$p(1 - p) < \frac{1}{4(1 + 1/\sqrt{n})^2}.$$

En el caso general la determinación de las expresiones exactas (funciones explícitas de X) para las estimaciones bayesianas y minimax no es siempre posible. Por eso es natural utilizar también el enfoque asintótico.

Antes de introducir las definiciones correspondientes, debemos recordar que las estimaciones bayesianas y minimas θ_Q^* y $\bar{\theta}^*$ han sido definidas por

las desigualdades

$$\mathbf{M}(\theta_Q^* - \theta)^2 - \mathbf{M}(\theta^* - \theta)^2 \leqslant 0,$$
$$\sup_{t\in\Gamma} \mathbf{M}_t(\bar{\theta}^* - t)^2 - \sup_{t\in\Gamma} \mathbf{M}_t(\theta^* - t)^2 \leqslant 0 \quad (9)$$

para cualquier estimación θ^*. No sería racional determinar el carácter bayesiano y minimax de las estimaciones, añadiendo simplemente a los primeros miembros el signo del paso límite ($\lim_{n\to\infty}$), ya que, por regla general, para las estimaciones a.n. de $\mathbf{M}_\theta(\theta^* - \theta)^2 \sim \sigma^2(\theta)/n$, los primeros miembros en (9) también convergerán hacia el cero. Por eso es natural examinar, digamos, la relación de los sumandos en (9). Teniendo en cuenta que más adelante se tratará principalmente de las estimaciones para las cuales $\mathbf{M}_\theta(\theta^* - \theta)^2$ tiene un orden de pequeñaz igual a $1/n$, se puede utilizar de un modo equivalente la definición siguiente.

Definición 3. La estimación θ_1^* se denomina *asintóticamente bayesiana o asintóticamente minimas*, si para cualquier otra estimación θ^* se cumple, respectivamente,

$$\lim_{n\to\infty}\sup\, [\mathbf{M}n(\theta_1^* - \theta) - \mathbf{M}n(\theta^* - \theta)^2] \leqslant 0,$$
$$\lim_{n\to\infty}\sup\, [\sup_{t\in\Gamma} \mathbf{M}_t n(\theta_1^* - t)^2 - \sup_{t\in\Gamma} \mathbf{M}_t n(\theta^* - t)^2] \leqslant 0.$$

Como veremos, la determinación de las estimaciones asintóticamente bayesianas y asintóticamente minimax es posible para suposiciones muy amplias.

En el *caso multidimensional* (cuando $\theta \in R^k$ es un vector) la propiedad 9) de la e.m.c., como hemos visto, se conserva, y la estimación

$$\theta_Q^* = \mathbf{M}(\theta/X)$$

minimizará

$$v(\theta^*) = \mathbf{M}(\theta^* - \theta)V(\theta^* - \theta)^T = \mathbf{M}\mathbf{M}_\theta(\theta^* - \theta)V(\theta^* - \theta)^T =$$
$$= \int \mathbf{M}_t(\theta^* - t)V(\theta^* - t)^T q(t)\lambda(dt)$$

para cualquier matriz V definida no negativamente o, que es lo mismo (véase el § 8), minimizará la desviación estándar $\theta^* - \theta$ promediada (con peso $q(t)$) en cualquier dirección $a \in R^k$.

Definición 4. La estimación θ_Q^* se llama *bayesiana* si para cualquier otra estimación θ^* y para cualquier matriz V definida no negativamente,

$$v(\theta_Q^*) \leqslant v(\theta^*).$$

La estimación θ_1^* se llama *asintóticamente bayesiana* si

$$\lim_{n\to\infty} \sup \,[nv(\theta_1^*) - nv(\theta_Q^*)] \leqslant 0.$$

Definición 5. La estimación $\bar{\theta}^*$ se denomina *minimax* si para cualquier otra estimación θ^* y para cualquier matriz V definida no negativamente,

$$\sup_{t\in\Gamma} \mathbf{M}_t(\bar{\theta}^* - t)V(\bar{\theta}^* - t)^T - \sup_{t\in\Gamma} \mathbf{M}_t(\theta^* - t)V(\theta^* - t)^T \leqslant 0.$$

La estimación θ_1^* se denomina *asintóticamente minimax* si

$$\lim_{n\to\infty} \sup \,[\sup_{t\in\Gamma} \mathbf{M}_t n(\theta_1^* - t)V(\theta_1^* - t)^T - \sup_{t\in\Gamma} \mathbf{M}_t n(\bar{\theta}^* - t)V(\bar{\theta}^* - t)^T] \leqslant 0.$$

Concluyendo este párrafo señalaremos una vez más que las designaciones $\mathbf{M}_\theta S$, $\mathbf{P}_\theta(A)$, $f_\theta(x)$ en el caso bayesiano pueden ser consideradas, si es necesario, desde un nuevo punto de vista: como esperanzas matemáticas, probabilidades y densidades condicionales respecto a θ, o sea, como $\mathbf{M}(S/\theta)$, $\mathbf{P}(A/\theta)$ y $f(x/\theta)$, respectivamente.

§ 12. Estadísticas suficientes

En el párrafo anterior hemos examinado la cuestión acerca de la construcción de dos tipos de estimaciones óptimas: bayesianas y minimáx. En este párrafo introduciremos el concepto de estadística suficiente, que nos permitirá construir estimaciones eficientes, o sea, otro tipo de estimaciones óptimas destacadas en el § 8.

La noción de estadística suficiente desempeña un papel importante en la estadística matemática en general y en la teoría de las estimaciones en particular.

Convengamos en designar las estadísticas, o sea, las funciones medibles arbitrarias (escalares o vectoriales) de X, con el símbolo $S = S(X)$.

Sea $X \in \mathbf{P}_\theta$, $\mathbf{P}_\theta \in \mathcal{P} = \{\mathbf{P}_\theta\}$. Examinemos la distribución $\mathbf{P}_\theta(X \in B/S)$, $B \in \mathfrak{B}^n_{\mathcal{X}}$ que es condicional respecto a la variable aleatoria S y que ha sido engendrada por la distribución $\mathbf{P}_\theta$ en $\mathcal{X}^n$.

Definición 1. La estadística $S = S(X)$ se llama *suficiente para el parámetro* θ, si existe la variante de la distribución condicional $\mathbf{P}_\theta(X \in B/S)$ que no depende de θ.,

Sabemos que $\mathbf{P}_\theta(X \in B/S)$ es, para cada B, la e.m.c. y, por consiguiente, existe una función $\mathbf{P}(B/s)$ de Borel en s para cada B, tal que

$$\mathbf{P}_\theta(X \in B/S) = \mathbf{P}(B/S).$$

Podemos considerar (véase el § 10) que $\mathbf{P}(B/s)$, como función de B, es la *distribución condicional de las probabilidades, a condición de que* $S = s$. Esta distribución puede interpretarse como la *distribución de* X en *la superficie* $S(x) = s$.

Pero si S es una estádística suficiente, entonces dicha distribución ¡*no depende de* θ! Esto significa que el conocimiento del lugar donde se encuentra el punto muestral X en la superficie $S(x) = s$ no nos comunica ninguna información complementaria acerca del parámetro θ. (Pues está claro que nadie se dedicará a determinar el parámetro desconocido en el ejemplo 1 de la Introducción, con ayuda del lanzamiento de una moneda, puesto que la distribución del número de "caras" o "cruces" con tal lanzamiento no depende de θ en absoluto).

Esta circunstancia importante significa, a su vez, que toda la información acerca del parámetro θ está contenida en el valor de la estadística S. De aquí precisamente procede su nombre: estadística suficiente. Hablando en términos generales, el conocimiento de $S(X)$ es *suficiente* para construir el parámetro θ, pero los demás datos contenidos en la muestra X son inútiles.

Ejemplo 1. Sea $X \in \Pi_\lambda$. Demostremos que la estadística $S = n\bar{x} = \sum_{i=1}^{n} x_i$ es suficiente para el parámetro de la ley de Poisson λ. Debemos convencernos de que la distribución de la posición del punto X en la superficie $\sum_{i=1}^{n} x_i = s$ (s es un número entero) no depende de λ. En vista de que $\mathbf{P}(X = \mathbf{x}, \sum x_i = s) = \mathbf{P}(X = \mathbf{x})$ cuando $\sum_{i=1}^{n} x_i = s$, entonces

$$\mathbf{P}(X = x/n\bar{x} = s) = \begin{cases} \dfrac{\mathbf{P}(x_1 = x_1, \ldots, x_n = x_n)}{\mathbf{P}(n\bar{x} = s)} & \text{si } \displaystyle\sum_{i=1}^{n} x_i = s, \\ 0 & \text{si } \displaystyle\sum_{i=1}^{n} x_i \neq s. \end{cases}$$

Como x_i son independientes, $\sum_{i=1}^{n} x_i \in \Pi_{n\lambda}$, el segundo miembro de (1) es igual a

$$\left(e^{-n\lambda}\frac{(n\lambda)^s}{s!}\right)^{-1}\prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = \frac{s!}{n^s\prod_{i=1}^{n} x_i!}.$$

Ahora bien, la distribución de X, que es condicional cuando $S = s$, coincide con la distribución polinomial $\mathbf{B}_p^s$ (véase el § 2) con n casos equiprobables (o sea, con el vector de probabilidades $p = (1/n, \ldots, 1/n)$) y con s pruebas independientes. Es evidente que la disribución no depende de λ, así que $S = n\bar{x}$ es una estadística suficiente para λ.

El concepto de estadística suficiente fue introducido en 1922 por Fisher. El siguiente teorema de Neyman — Fisher lleva el nombre de teorema de factorización y establece un criterio elemental de existencia de la estadística suficiente.

Supongamos que ha sido cumplida la condición (A_μ) de existencia de la densidad $f_\theta(x) = \frac{d\mathbf{P}_\theta}{d\mu}(x)$.

Teorema 1. *Para que S sea una estadística suficiente para θ, es necesario y suficiente que la función de verosimilitud* $f_\theta(x) = \prod_{i=1}^{n} f_\theta(x_i)$ *sea representable en la forma*

$$f_\theta(x) = \psi(S(x), \theta)h(x) \quad \text{c.s.}[\mu^n], \tag{2}$$

donde cada una de las funciones $\psi \geqslant 0$ y $h \geqslant 0$ *depende sólo de sus propios argumentos,* $\psi(s, \theta)$ *es medible en s, y* $h(x)$*, en x.*

Por supuesto que la representación (2) no es unívoca. Sus componentes han sido determinados con una exactitud de hasta una función positiva arbitraria de $S(\bar{x})$.

En el ejemplo anteriormente examinado, con la distribución de Poisson,

$$f_\lambda(x) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\lambda^{n\bar{x}} \prod_{i=1}^{n} \frac{1}{x_i!}, \quad n\bar{x} = \sum_{i=1}^{n} x_i,$$

así que podemos, para $S = n\bar{x}$, poner

$$\psi(S, \lambda) = e^{-n\lambda}\lambda^S h, \quad h(\bar{x}) = \prod_{i=1}^{n} \frac{1}{x_i!}.$$

De aquí, en virtud del teorema 1, resultará que $S = n\bar{x}$ es una estadística suficiente.

La demostración del teorema 1 aquí sólo se da para dos casos particulares más importantes: para el caso discreto y para el caso "suave". En el caso general, la demostración del teorema de Neyman — Fisher se da en el Suplemento IV.

En el *caso discreto*, μ es la medida de cálculo en el conjunto numerable $\mathscr{X}$ de los posibles valores de x_1 y, por lo tanto, $f_\theta(x) = \mathbf{P}_\theta(x_1 = x)$, $x \in \mathscr{X}$. Supongamos que al principio ha sido cumplida (2). Entonces, para el punto registrado $x \in \mathscr{X}^n$,

$$\mathbf{P}_\theta(X = x/S(X) = S(x)) = \frac{\mathbf{P}_\theta(X = x,\ S(X) = S(x))}{\mathbf{P}_\theta(S(X) = S(x))}. \tag{3}$$

Como $\{X = x, S(X) = S(x)\} = \{X = x\}$, el segundo miembro de (3) es igual a

$$\frac{\mathbf{P}_\theta(X = x)}{\mathbf{P}_\theta(S(X) = S(x))} = \frac{f_\theta(x)}{\sum\limits_{Y: S(y) = S(x)} f_\theta(y)} =$$

$$= \frac{\psi(S(x), \theta)\, h(x)}{\sum\limits_{y: S(y) = S(x)} \psi(S(x), \theta) h(y)} = \frac{h(x)}{\sum\limits_{y: S(y) = S(x)} h(y)}.$$

Ahora bien, $\mathbf{P}_\theta(X = x/S(X) = S(x)$ no depende de θ.

Al contrario, si el primer miembro de (3) no depende de θ, entonces, designándolo por $h(x)$, de (3) obtenemos $\mathbf{P}_\theta(X = x) = f_\theta(x) = \mathbf{P}_\theta(X = x; S(X) = S(x)) = h(x)\ \mathbf{P}_\theta(S(X) = S(x))$, donde $\mathbf{P}_\theta(S(X) = S(x)) = \psi(S(x), \theta)$ depende solamente de $S(x)$ y de θ. ◁

De un modo algo más complicado el teorema 1 también se demuestra en otro importante caso particular, o sea, en el caso "*suave*" cuando μ es la medida de Lebesgue en R, y la estadística $S(X)$ se supone que es función suave de X, es decir, una función tal que existe la sustitución de las variables $y_1 = S(x)$, $y_2 = y_2(x)$, ..., $y_n = y_n(x)$, resoluble respecto a $x_i = x_i(y_1, \ldots, y_n)$, con un jacobiano distinto del cero $J = \left|\frac{\partial x_i}{\partial y_j}\right| \neq 0$. En este caso, como es sabido de las fórmulas del análisis clásico sobre la sustitución de la variable en la integral, la densidad de la variable aleatoria $Y = (S(X), y_2(X)), \ldots, y_n(X))$ será igual a

$$g_\theta(y) = f_\theta(x)|J|,\ y = (y_1, \ldots, y_n).$$

La densidad de la variable aleatoria $y_1(X) = S(X)$ será igual a

$$g_\theta^{(1)}(y_1) = \int\limits_{R^{n-1}} g_\theta(y) dy_2 \ldots dy_n = \int\limits_{R^{n-1}} f_\theta(x)|J| dy_2 \ldots dy_n,$$

y la condicional de Y, a condición de que $S(X) = s$, será, por consiguiente, determinada por la relación

$$\varphi(y/s) = \frac{g_\theta(y)}{g_\theta^{(1)}(s)} = \frac{f_\theta(x)|J|}{g_\theta^{(1)}(s)} \text{ para } y_1 = s.$$

Después de estas observaciones preliminares, la demostración del teorema 1 para el caso "suave" se desarrolla al igual que para el caso discreto. En efecto, si se ha cumplido (2), entonces

$$\varphi(y/s) = \frac{\psi(s, \theta) h(x)|J|}{\int\limits_{R^{n-1}} \psi(s, \theta) h(x)|J| dy_2 \ldots dy_n}.$$

En esta relación, $\psi(s, \theta)$ se reduce. Esto significa que la distribución de Y, condicional a condición de que $S(X) = s$, y, por lo tanto, también la distribución de X no depende de θ,

Al contrario, si $\varphi(y/s)$ no depende de θ, entonces

$$f_\theta(x) = \frac{\varphi(y/s)g_\theta^{(1)}(s)}{|J|} \text{ cuando } s = S(x).$$

Esto significa que (2) se cumple cuando $\psi(s, \theta) = g_\theta^{(1)}(s)$, $h(x) = \varphi(y/s)/|J|$. ◁

Ejemplo 2. Sea $X \in \Phi_{\alpha,\sigma^2}$. Aquí el parámetro $\theta = (\alpha, \sigma^2)$ es bidimensional. Tenemos

$$f_\theta(X) = \prod_{i=1}^{n} \frac{-1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\alpha)^2}{2\sigma^2}} = \sigma^{-n}(2\pi)^{-n/2}\exp\left\{\frac{\sum(x_i-\alpha)^2}{2\sigma^2}\right\} =$$

$$= \sigma^{-n}\exp\left\{-\frac{\sum x_i^2 - 2\alpha n\bar{x} + n\alpha^2}{2\sigma^2}\right\}(2\pi)^{-n/2}.$$

Poniendo $S = (S_1, S_2)$, $S_1 = n\bar{x}$, $S_2 = \sum_{i=1}^{n} x_i^2$, obtenemos la representación (2), donde

$$\psi(S, \theta) = \sigma^{-n}\exp\left\{-\frac{S_2 - 2\alpha S_1 + \alpha^2}{2\sigma^2}\right\}, \quad h(X) = (2\pi)^{-n/2}.$$

Aquí podríamos, desde luego, atribuir el factor $(2\pi)^{-n/2}$ también a la función ψ, poniendo $h(X) = 1$.

Ahora bien, hemos obtenido que la estadística (S_1, S_2) es una estadística vectorial suficiente para (α, σ^2). De toda la información contenida en la muestra nos es suficiente saber $\bar{x}$ y $\sum x_i^2$.

Proponemos al lector hallar las estadísticas suficientes para todas las familias de distribuciones citadas en el § 2.

Concentraremos la atención tan sólo en una de estas familias.

Ejemplo 3. Sea $X \in U_{0,\theta}$. Aquí la condición (A_μ) se cumple con respecto a la medida de Lebesgue y

$$f_\theta(X) = \begin{cases} \theta^{-n} \text{ si } 0 \leqslant x_i \leqslant \theta \text{ cuando todos } i = 1, \ldots, n \\ 0 \text{ en el caso contrario.} \end{cases}$$

Sea $x_{(1)} = \text{mín } x_i$, $x_{(n)} = \text{máx } x_i$. Entonces, como hemos visto en el ejemplo 6.5, la función $f_\theta(X)$ puede ser escrita en forma de $f_\theta(X) = \psi(x_{(n)}, \theta)h(X)$, donde

$$h(X) = \begin{cases} 1 \text{ si } x_{(1)} \geqslant 0, \\ 0 \text{ en el caso contrario,} \end{cases}$$

$$\psi(s, \theta) = \begin{cases} \theta^{-n} & \text{para } s \leqslant \theta, \\ 0 & \text{en el caso contrario.} \end{cases}$$

Esto significa que $S(X) = x_{(n)}$ es una estadística suficiente para θ.

Análogamente el lector puede convencerse de que para la muestra $X \in \in U_{\theta, 1+\theta}$, como estadística suficiente para el parámetro θ, sirve la estadística bidimensional $S(X) = x_{(1)}, x_{(n)})$. Asimismo será la estadística suficiente para el parámetro bidimensional $\theta = (a, b)$ cuando la muestra ha sido extraída de la distribución $U_{a,b}$.

Citaremos dos corolarios del teorema 1.

Corolario 1. *Si S es una estadística suficiente para θ, la estimación de verosimilitud máxima depende únicamente de S.*

Mejor dicho, la e.v.m. $\hat{\theta}^*$ no depende de X cuando se ha registrado $S(X)$.

Este corolario es evidente, ya que la e.v.m. es un valor de θ para el cual se alcanza el máximo de $f_\theta(X) = \psi(S(X), \theta)h(X)$ o bien, que es lo mismo, el máximo de $\psi(S(X), \theta)$.

Corolario 2. *Si S es una estadística suficiente y la función φ es tal que la aplicación $u = \varphi(v)$ es biunívoca y medible en ambas direcciones, entonces $S_1 = \varphi(S)$ también será una estadística suficiente.*

Este corolario también es evidente, puesto que $\psi(S, \theta)$ en (2) puede escribirse en forma de $\psi(\varphi^{-1}(S_1), \theta) = \psi_1(S_1, \theta)$.

También es válido un criterio más de suficiencia de la estadística S.

Teorema 2. *La estadística S es suficiente para θ si y sólo si para toda distribución a priori* **Q** *del parámetro θ la distribución a posteriori* $\mathbf{Q}_X$ *depende de X tan sólo a tráves de S(X) (o sea, permanece invariable en la superficie de S(X) = S).*

Demostración. Supongamos que S es una estadística suficiente y que $q(t)$ es la densidad **Q** respecto a cualquier medida λ. Entonces, la densidad a posteriori $q(t/X)$ respecto a dicha medida, según la fórmula de Bayes será igual a

$$q(t/X) = \frac{f_t(X)q(t)}{\int f_u(X)q(u)\lambda(du)} = \frac{\psi(S(X), t)q(t)}{\int \psi(S(X), u)q(u)\lambda(du)}.$$

Demostremos ahora la afirmación inversa del teorema. Escojamos una distribución a priori de modo que $q(t) > 0$ en todas partes sobre Θ y para todos t

$$f_t(X) = \frac{q(t/X)f(X)}{q(t)}, \quad f(X) = \int f_u(X)q(u)\lambda(du).$$

Si $q(t/X) = g(t, S(X))$, entonces, poniendo $\psi(s, t) = g(t, s)/q(t)$, $h(X) = = f(X)$, obtenemos la representación (2). ◁

Corolario 3. *Si S es una estadística suficiente, todas las estimaciones*

bayesianas y las estimaciones minimax definidas con ayuda del teorema 11.2 dependen únicamente de S.

En adelante obtendremos muchas otras confirmaciones de que la estadística suficiente S contiene la información completa acerca de θ.

§ 13*. Estadísticas suficientes mínimas

Examinemos ahora la cuestión acerca de la elección de las características suficientes. Claro está que el número de éstas puede ser muy grande. Por ejemplo, la estadística $S(X) \equiv X$ siempre es evidentemente suficiente. La misma se llama estadística suficiente *trivial*. Sin embargo, estamos interesados (posteriormente será aclarado el porqué) en estadísticas más "económicas". Resulta que no siempre, ni micho menos, se pueden construir estadísticas suficientes que sean mucho más "económicas" que la estadística suficiente trivial. Volveremos a esta cuestión después que determinemos más exactamente los conceptos relacionados con la "economía" de las características suficientes. Para esto, introduzcamos en el conjunto de todas las características suficientes (para cierto parámetro θ), un orden parcial.

Definición 1. Diremos que la característica S_1 *está subordinada* a S_2 si S_1 es una función medible de S_2:$S_1 = \varphi(S_2)$.

Esta relación significa precisamente que S_1 es más "económica" que S_2.

Definición 2. Si S_1 está subordinada a S_2, y S_2 está subordinada a S_1, las estadísticas S_1 y S_2 se denominan *equivalentes*.

Evidentemente, S_1 es equivalente a S_2 si y sólo si $S_1 = \varphi(S_2)$ y φ es una aplicación biunívoca medible en ambas direcciones.

Definición 3. La estadística suficiente S_0 se denomina *mínima* si está subordinada a cualquier otra estadística suficiente S.

La estadística suficiente mínima es la más económica. Si hemos construido la estadística suficiente mínima S, entonces, siempre que se conserve la propiedad de suficiencia, será imposible la reducción ulterior de los datos en comparación con S. Los demás datos contenidos en la muestra pueden considerarse como engendrados por cierto mecanismo aleatorio no dependiente de θ, y ellos no proporcionan ninguna información acerca de θ.

Los conceptos introducidos, al igual que el concepto inicial de estadística suficiente, pueden exponerse, de forma ligeramente generalizada, en el lenguaje de las σ-álgebras, que en una serie de casos resulta más cómodo y evidente. Al principio —en la definición 1 del párrafo precedente— la distribución condicional $\mathbf{P}_\theta(X \in B/S)$ se puede sustituir por la distribución condicional $\mathbf{P}_\theta(X \in B/\mathfrak{U})$ respecto a la σ-subálgebra $\mathfrak{U} \subset \mathfrak{B}_{\mathscr{X}}$ y la $\mathfrak{U}$ *σ-álgebra se puede llamar suficiente* si existe cierta variante $\mathbf{P}_\theta(X \in B/\mathfrak{U})$ que no depende de θ.

Con tal enfoque, el teorema de factorización se conserva si la función $\psi(S(X), \theta)$ es sustituida por la función $\psi(X, \theta)$ $\mathfrak{U}$-medible en X. La demostración de este teorema, expuesta en el Suplemento IV, prácticamente no se diferencia de la anterior.

La estadística suficiente ahora puede ser definida como una estadística S para la cual la σ-álgebra de $\sigma(S)$ será suficiente.

En el lenguaje de las σ—álgebras, la subordinación de las características suficientes (véase la definición 1) no exige que se introduzcan conceptos complementarios y coincide simplemente con el encaje de las σ-álgebras: S_1 está subordinada a S_2 si $\sigma(S_1) \subset \sigma(S_2)$. Ahora bien, S_1 es más económica que S_2 si la σ-álgebra de $\sigma(S_1)$ es más pobre que $\sigma(S_2)$. La equivalencia de S_1 y S_2 significa que $\sigma(S_1) = \sigma(S_2)$.

La *σ-álgebra suficiente mínima* de $\mathfrak{U}_0$ se define como una σ-álgebra que se encaja en cualquier σ-álgebra suficiente.

La *σ-álgebra suficiente mínima existe siempre.* Para convencerse de ello señalaremos previamente que, en virtud del teorema 2 del Suplemento IV, existe una distribución $\mathbf{Q}$ en Θ (además, discreta), tal que todas $\mathbf{P}_\theta$ son absolutamente continuas respecto a la distribución $\mathbf{P}_Q = \int \mathbf{P}_t \mathbf{Q}(dt)$.

Esto significa que $f_Q(X) = \int f_t(X)\mathbf{Q}(dt) > 0$ para todas X, o que de la igualdad $f_Q(X) = 0$ resulta $f_\theta(X) = 0$ para todos θ. En este caso se dice que $\mathbf{P}_Q$ *domina* la familia $\{\mathbf{P}_\theta\}$, así que podríamos adoptar $\mathbf{P}_Q$ como medida de μ. La densidad de la distribución $\mathbf{P}_\theta$ respecto a esta medida es igual a

$$\frac{d\mathbf{P}_\theta}{d\mathbf{P}_Q}(x) = \frac{f_\theta(x)}{f_Q(x)} \equiv r(x, \theta).$$

Está claro (compárese con el teorema 12.2) que si S es una estadística suficiente, $r(x, \theta)$ depende de x sólo a través de $S(x)$.

Teorema 1. *La σ-álgebra de $\mathfrak{U}_0 = \sigma(r(X, \theta); \theta \in \Theta)$ engendrada por las variables aleatorias $r(X, \theta) = f_\theta(X)/f_Q(X)$ para diferentes $\theta \in \Theta$, es una σ-álgebra suficiente mínima.*

La demostración del teorema es muy simple. La suficiencia de $\mathfrak{U}_0$ resulta del teorema de factorización y del hecho de que

$$f_\theta(X) = r(X, \theta)f_Q(X), \tag{1}$$

donde $f_Q(X)$ no depende de θ, y $r(X, \theta)$ es medible respecto a $\mathfrak{U}_0$.

Sea ahora $\mathfrak{U}$ cualquier σ-álgebra suficiente. Entonces $f_\theta(X) = \psi(X, \theta)h(X)$, donde la función $\psi(X, \theta)$ es $\mathfrak{U}$-medible. Examinemos la σ-álgebra de $\mathfrak{U}_\psi = \sigma(\psi(X, \theta), \theta \in \Theta) \subset \mathfrak{U}$. De la definición $r(X, \theta)$ se deduce que

$$r(X, \theta) = \frac{\psi(X, \theta)}{\int \psi(X, t)\mathbf{Q}(dt)}$$

y, por lo tanto, $\mathfrak{U}_0 \subset \mathfrak{U}_\psi \subset \mathfrak{U}$. ◁

Con este teorema y con el teorema 12.2 está estrechamente ligada otra afirmación útil. Examinemos el planteamiento bayesiano del problema cuando θ es una variable aleatoria con la distribución a priori $\mathbf{Q}$. Sea $q(t) > 0$ la densidad de esta distribución con respecto a la medida conveniente λ en Θ. Entonces la densidad a posteriori será igual a

$$q(t/X) = \frac{f_t(X)q(t)}{f_Q(X)} = r(X,\ t)q(t),$$

y, por consiguiente, la σ-álgebra suficiente mínima de $\mathfrak{U}_0$ puede considerarse como engendrada por la distribución a posteriori:

$$\mathfrak{U}_0 = \sigma(q(t/X);\ t \in \Theta).$$

Por regla general, la determinación de las distribuciones $\mathbf{Q}$ y $\mathbf{P}_Q$ que figuran en el teorema 1 no es difícil. Por ejemplo, si el portador $N_{P\theta}$ de la distribución $\mathbf{P}_\theta$ no depende de θ, lo que tiene lugar para la mayoría de las distribuciones citadas en el § 2, se puede tomar $\mathbf{P}_Q = \mathbf{P}_{\theta_0}$ para cualquier $\theta_0 \in \Theta$.

Así pues, disponemos del teorema de existencia y del método eficaz para la construcción de las σ-álgebras suficientes mínimas [*)].

No obstante, las más de las veces para nosotros será más cómodo examinar las estadísticas. El fin principal de este párrafo consiste en determinar las estadísticas suficientes mínimas.

Ante todo, ¿de qué modo podemos comprobar que la estadística suficiente dada S_0 es mínima?

Una de las posibilidades consiste en la utilización del teorema 1. Si $\sigma(S_0)$ coincide con la σ-álgebra engendrada por $f_\theta(X)/f_Q(X)$, entonces S_0 es la estadística suficiente mínima.

Ejemplo 1. Hemos visto que la estadística $S = n\bar{x}$ es suficiente para el parámetro λ de la distribución de Poisson Π_λ. Ella será la estadística suficiente mínima, ya que $\sigma(S)$ coincide, evidentemente, con la σ-álgebra engendrada por $f_\lambda(X)/f_{\lambda_1}(X) = e^{n(\lambda_1 - \lambda)}(\lambda/\lambda_1)^S$ (aquí hemos tomado la distribución $\mathbf{Q}$ concentrada en el punto λ_1).

Ejemplo 2. Sea $X \in \mathbf{U}_{0,\theta}$. Entonces la estadística $S = x_{(n)} = \max x_i$ es la estadística suficiente mínima. En efecto, tomemos en calidad de $\mathbf{Q}$ cualquier distribución sobre $[0, \infty)$ con desindad $q(t) > 0$ para todos $t > 0$. Entonces

$$f_\theta(X) = \begin{cases} \theta^{-n}, & \theta \geqslant S, \\ 0, & \Theta < S, \end{cases}$$

[*)] La existencia de la σ-álgebra suficiente mínima de $\mathfrak{U}_0$ también se puede establecer de otra manera, demostrando que $\mathfrak{U}_0$ es la intersección de todas las σ-álgebras suficientes completadas.

$$f_Q(X) = \int_0^\infty f_t(X)q(t)dt = \int_S^\infty t^{-n}q(t)dt > 0$$

para todas X. En este caso $S = \sup\{\theta; f_\theta(X)/f_Q(X) = 0\}$, lo cual significa que S es medible respecto a la σ-álgebra mínima de $\mathfrak{U}_0$, $\sigma(S) \subset \mathfrak{U}_0$ y que, por lo tanto, S es la estadística suficiente mínima.

Podemos indicar otro método de determinar las estadísticas suficientes mínimas, el cual también está relacionado con la función de verosimilitud. En efecto, toda estadística y, en particular, la estadística suficiente S engendra la partición del espacio muestral en clases de equivalencia, o sea, en conjuntos de los puntos x con iguales valores de $S(x)$.

Si S_1 está subordinada a S_2, o sea, $S_1 = \varphi(S_2)$, es evidente que para S_1 la partición será más grande, ya que las clases de equivalencia para S_2 se contienen en las de equivalencia para S_1. Ahora bien, a la estadística suficiente mínima le corresponde la "mayor" partición entre las particiones engendradas por las estadísticas suficientes.

Se pueden examinar simplemente las particiones del espacio en clases de equivalencia sin relacionarlas directamente con las estadísticas. Designemos por $D(x)$ la clase de equivalencia que contiene el punto x. Cada clase se define unívocamente por un punto cualquiera. *Llamaremos suficiente la partición en clases D si*

$$f_\theta(x) = \varphi(x, \theta)h(x), \tag{2}$$

donde $\varphi(x, \theta) = \varphi(x_\theta, \theta)$ *es constante para* $x \in D(x_0)$ (o sea, $\varphi(x, \theta) =$ const dentro de la clase de equivalencia). Si las clases $D(x)$ son definidas por las relaciones $S(x) = s$, del teorema 11.1 se desprende directamente que la estadística $S(x)$ es suficiente si y sólo si la partición en clases D es suficiente.

Examinemos ahora la partición construida del modo siguiente: tomemos el punto x_0 y declaremos que x pertenece a la clase $D(x_0)$ si la relación

$$\frac{f_\theta(x)}{f_\theta(x_0)} = h(x, x_0) \tag{3}$$

no depende de θ. Es evidente que con tal construcción, $D(x_1) = D(x_2) = D(x_0)$ si $x_1 \in D(x_0)$, $x_2 \in D(x_0)$, así que la regla (3) engendra la partición de todo el espacio en clases disjuntas.

Esta partición corresponde a la engendrada por la estadística suficiente mínima S.

En efecto, sea S la estadística suficiente mínima. Tomemos un punto arbitrario x_0. Entonces sobre la superficie $S(x) = S(x_0)$, la relación $f_\theta(x)/f_\theta(x_0)$ es igual a $h(x)/h(x_0)$ y, por consiguiente, no depende de Θ. Así pues, la partición en clases D es no menos grande que la partición para S.

Por otro lado, esta partición es suficiente. Efectivamente, podemos hacer que a cada superficie D le corresponda un punto cualquiera x_D de ella, a partir del cual la misma será definida unívocamente. Examinemos la función $x_0(x)$ que se define según la relación $x_0(x) = x_D$ si $x \in D$. Entonces, en virtud de (3), cuando $x \in D$,

$$f_\theta(x) = f_\theta(x_D)h(x, x_D) = f_\theta(x_0(x))h(x_1, x_0(x)), \tag{4}$$

que significa el cumplimiento de (2).

Los planteamientos efectuados no han sido del todo estrictos, ya que no los hemos relacionado con la cuestión acerca de la mensurabilidad de las funciones que forman parte de (4).

Lo dicho se puede resumir del modo siguiente. *Supongamos que se da una estadística* $S(X)$ *tal que* $S(x) = S(x_0)$ *si y sólo si la relación* (3) *no depende de* θ. *En este caso S es la estadística suficiente mínima.*

A distinción de los enfoques relacionados con el teorema 1, donde fueron examinadas las relaciones $f_\theta(x)/f_Q(x)$ o bien $f_\theta(x)/f_{\theta_1}(x)$ para diferentes θ y θ_1 (denominadas con frecuen-

cia relaciones de verosimilitud), la regla enunciada más arriba utiliza la relación $f_\theta(x)/f_\theta(x_0)$ para iguales valores del parámetro θ. En el ejemplo 1, por ejemplo, la relación

$$f_\lambda(x)/f_\lambda(x_0) = \prod \lambda^{x_i - x_{i0}} x_{i0}!/x_i! = \lambda^{n(\bar{x} - \bar{x}_0)} \prod x_{i0}!/x_i!$$

no dependerá de λ si y sólo si $\bar{x} = \bar{x}_0 = \frac{1}{n} \sum_{i=1}^{n} x_{i0}$, donde x_{i0} son las coordenadas del vector x_0. Esto es suficiente para sacar la conclusión de que $S(x) = \bar{x}$ es la estadística suficiente mínima.

Valiéndonos de la regla propuesta, examinemos ahora un ejemplo cuando no existen estadísticas suficientes " económicas". Antes que nada señalaremos que la serie variacional $S_V = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$, construida según la muestra X, es siempre, evidentemente, la estadística suficiente, ya que $f_\theta(X) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{k=1}^{n} f_\theta(x_{(k)})$. Esta estadística es "un poco más económica" que la propia muestra x. De aquí, en particular, se deduce que cualquier estadística suficiente mínima es invariante con respecto a la permutación de las coordenadas x_i en la muestra X.

Si la densidad $f_\theta(x)$ es simétrica, o sea, $f_\theta(-x) = f_\theta(x)$ para todos θ, es evidente que existirá una estadística suficiente, un poco más "económica", que representa la población $(x_1^2, \ldots, x_n^2)$ ordenada en función de su crecimiento y que designaremos por S_{V^2}.

Ejemplo 3. Si $X \in \mathbf{K}_{0,\sigma}$, o sea, si x_i tiene densidad de distribución de Cauchy con parámetro $\theta = \sigma$,

$$k_{0,\sigma}(x) = \frac{\sigma}{\pi(x^2 + \sigma^2)},$$

la estadística S_{V^2} será la estadística suficiente mínima.

En efecto, en este caso

$$f_\sigma(x) = \left(\frac{\sigma}{\pi}\right)^n \prod_{i=1}^{n} (x_i^2 + \sigma^2)^{-1},$$

así que

$$\frac{f_\sigma(x)}{f_\sigma(x_0)} = \prod_{i=1}^{n} \frac{x_{i0}^2 + \sigma^2}{x_i^2 + \sigma^2} \tag{5}$$

es la relación de dos polinomios de σ^2, la cual no depende de σ si y sólo si los coeficientes de las potencias correspondientes de σ^2 coinciden en el numerador y el denominador. Esto, a su vez, tiene lugar si y sólo si los conjuntos de "ceros" $\{-x_{i0}^2\}$ y $\{-x_i^2\}$ coinciden. Con otras palabras, para que (5) sea independiente de σ es necesario y suficiente que el punto $x^2 = (x_1^2, \ldots, x_n^2)$ tenga coordenadas que se distingan de las de x_0^2 tan sólo por la permutación de sus lugares. Esto precisamente significa que S_{V^2} es una estadística suficiente mínima.

De manera completamente análoga se puede demostrar que S_V es una estadística suficiente mínima para el parámetro α y, por lo tanto, para el parámetro $\theta = (\alpha, \sigma)$ de la distribución $\mathbf{K}_{\alpha,\sigma}$.

Otro ejemplo, en el que S_V será una estadística sifuciente minima, se obtiene si se examina la familia

$$\mathbf{P}_{\alpha,\theta_1,\theta_2} = \alpha \mathbf{P}_{\theta_1} + (1 - \alpha)\mathbf{P}_{\theta_2}, \qquad \alpha \in [0, 1],$$

donde $\{\mathbf{P}_\theta\}$ es una familia exponencial (véase § 15, en calidad de $\mathbf{P}_\theta$ se puede tomar la distri-

bución normal o la distribución de Poisson) y donde al menos uno de los parámetros α, θ_1, θ_2 se desconoce.

Ahora demostremos un teorema que indica un método "estructural" simple de determinación de las estadísticas suficientes mínimas.

Para simplificar la exposición examinemos el caso del parámetro unidimensional θ.

Teorema 2. *Supongamos que la función de verosimilitud $f_\theta(x)$, para todas x como función de θ, es continua a la derecha (o a la izquierda). Entonces, si la estimación de v.m. $\hat{\theta}^*$ es única y la misma es una estadística suficiente, entonces $\hat{\theta}^*$ será la estadística suficiente mínima.*

Demostración. Sea S una estadística suficiente arbitraria. Demostraremos el teorema si mostramos que $\hat{\theta}^*$ es medible respecto a $\sigma(S)$ y, por lo tanto, $\hat{\theta}^*$ está subordinada a S.

En virtud del teorema de factorización,

$$f_\theta(x) = \psi(S(x), \theta)h(x) \text{ c.s.}[\mu^n], \tag{6}$$

donde $h(x)$ es la función medible en x, y $\psi(s, t)$ es continua (a la derecha o a la izquierda) en t y medible en s. Como $\mathbf{P}_\theta$ no variará si la densidad $f_\theta(x)$ cambia en el conjunto de la μ^n-medida 0, podemos considerar que (6) es válida para todos x.

En virtud de (6), el punto del máximo absoluto de $f_\theta(x)$ también es el punto del máximo absoluto para $\psi(S(x), \theta)$. Por eso, en virtud de la unicidad de $\hat{\theta}^*$,

$$\{\hat{\theta}^* < t\} = \{\sup_{\theta < t} \psi(S(X), \theta) > \sup_{\theta \geq t} \psi(S(X), \theta).$$

En vista de que $\psi(S(X), \theta)$, para cada $S(X)$, es continua en θ a la derecha (o a la izquierda), existe un conjunto numerable, denso en todas las partes, $\Theta_c = \{\theta_j\}_{j=1}^\infty \subset \Theta$ (igual para todos los $S(X)$) tal que

$$\sup_{\theta < t} \psi(S(X), \theta) = \sup_{\substack{\theta_j < t \\ \theta_j \in \Theta_c}} \psi(S(X), \theta_j). \tag{7}$$

Esa misma relación será válida para la región de $\theta \geqslant t$. Como $\psi(S(X), \theta_j)$ son medibles respecto a $\sigma(S)$, en virtud de (7), los valores de $\sup_{\theta < t} \psi(S, \theta)$ y $\sup_{\theta \geq t} \psi(S, \theta)$ serán variables aleatorias también medibles con respecto a $\sigma(S)$. Por consiguiente, $\{\theta^* < t\} \in \sigma(S)$, y el teorema ya está demostrado. ◅

En la condición de la afirmación citada, la condición de suficiencia de la e.v.m. $\hat{\theta}^*$ es esencial, puesto que la estimación de verosimilitud máxima $\hat{\theta}^*$, como tal, no es obligatoriamente una estimación suficiente. Es fácil

obtener un ejemplo respectivo examinando cualquier familia de distribuciones $\{\mathbf{P}_\theta\}$, con parámetro escalar θ y con estadística suficiente mínima vectorial S(cuya dimensión es mayor que 1). En este caso la estimación de verosimilitud máxima $\hat{\theta}^*$ también será escalar, así que la σ-álgebra de $\sigma(S)$ será más rica que $\sigma(\hat{\theta}^*)$ y, por lo tanto, la inclusión de $\sigma(S) \subset \sigma(\hat{\theta}^*)$, que se desprende de la minimalidad de S y de la suficiencia de $\hat{\theta}^*$, es imposible.

Ejemplo 4. Sea $X \in \mathrm{U}_{\theta,1+\theta}$, $\Theta = R$. Entonces, como hemos visto en el ejemplo 6.4,

$$f_\theta(X) = \begin{cases} 1 & \text{para } \theta \leqslant x_{(1)} \leqslant x_{(n)} \leqslant 1 + \theta, \\ 0 & \text{en el caso contrario,} \end{cases}$$

así que $f_\theta(X)$ depende de X solamente a través de $x_{(1)}$ y $x_{(n)}$. Esto significa que $S = (x_{(1)}, x_{(n)})$ es una estadística suficiente. Ni una de las magnitudes $x_{(1)}$, $x_{(n)}$ por separado es una estadística suficiente. Eso lo demuestran las relaciones siguientes:

$$\mathbf{P}(x_{(1)} \geqslant u,\ x_{(n)} < v) = \prod_{i=1}^{n} \mathbf{P}(x_i \in [u,\ v)) =$$
$$= (v - u)^n \text{ cuando } u \geqslant \theta,\ v \leqslant 1 + \theta,\ v > u.$$

Por consiguiente, la densidad compatible de distribución de $(x_{(1)}, x_{(n)})$ será igual a

$$g(u,\ v) = \begin{cases} n(n-1)(v-u)^{n-2} & \text{cuando } u \geqslant \theta,\ v \leqslant 1 + \theta,\ v > u, \\ 0 & \text{en los demás casos.} \end{cases}$$

Seguidamente, $\mathbf{P}(x_{(1)} \geqslant u) = (1 + \theta - u)^n$ cuando $\theta \leqslant u \leqslant 1 + \theta$, así que la densidad de $x_{(1)}$ es igual a

$$g(u) = n(1 + \theta - u)^{n-1} \text{ cuando } \theta \leqslant u \leqslant 1 + \theta.$$

De aquí ya es fácil obtener que la densidad condicional $g(v/u)$ de la magnitud $x_{(n)}$, a condición de $x_{(1)} = u$ (y, por lo tanto, también la distribución condicional correspondiente), dependerá de θ. Esto significa que $x_{(1)}$ (al igual que $x_{(n)}$) por separado no son estadísticas suficientes. Como en calidad de e.v.m. $\hat{\theta}^*$ podemos tomar $\hat{\theta}^* = x_{(1)}$ (véase el ejemplo 6.4, por lo tanto, hemos demostrado que para la familia $\mathrm{U}_{\theta,1+\theta}$, la e.v.m. $\hat{\theta}^*$ no es una estadística suficiente.

Mediante el teorema 1, el lector puede convencerse personalmente de que $S = (x_{(1)}, x_{(n)})$ es una estadística suficiente mínima para $\mathrm{U}_{\theta,1+\theta}$.

La condición de suficiencia de $\hat{\theta}^*$ en el teorema 2 será cumplida automáticamente si suponemos que existe una estadística suficiente escalar (para un θ unidimensional) S_0, para la cual la función φ en la igualdad $\hat{\theta}^* = \varphi(S_0)$ será biunívoca (o sea, $\hat{\theta}^*$ y S_θ serán equivalentes).

§ 14. Construcción de estimaciones eficientes por medio de estadísticas suficientes. Estadísticas completas

Definición 1. La estimación θ^* se denomina *suficiente* si es una estadística suficiente.

1. Caso unidimensional. Supondremos aquí que θ es un parámetro escalar. Sea K_b la clase de todas las estimaciones θ^* con desplazamiento $b(\theta)$, así que $\theta^* \in K_b$ si $a(\theta) = \mathbf{M}_\theta\theta^* = \theta + b(\theta)$. Para $\theta^* \in K_b$ tenemos

$$\mathbf{M}_\theta(\theta^* - \theta)^2 = \mathbf{M}_\theta(\theta^* - a(\theta))^2 + (a(\theta) - \theta)^2 = \mathbf{D}_\theta\theta^* + b^2(\theta).$$

En este párrafo omitiremos, a veces, el índice θ de los símbolos $\mathbf{M}_\theta$, $\mathbf{D}_\theta$.

La siguiente afirmación fue obtenida independientemente por Blackwell, Rao y Kolmogórov.

Teorema 1. *Sea S una estadística suficiente, $\theta^* \in K_b$. Entonces la función $\theta_S^* = \mathbf{M}_\theta(\theta^*/S)$ es una estimación que posee las siguientes propiedades:*

1) $\theta_S^* \in K_b$,

2) *θ_S^* depende de la muestra tan sólo a través de $S(X)$,*

3) $\mathbf{M}_\theta(\theta_S^* - \theta)^2 \leqslant \mathbf{M}_\theta(\theta^* - \theta)^2$ *para todos θ.*

La última desigualdad se transforma en igualdad tan sólo si $\theta^* = \theta_S^*$ c.d. respecto a $\mathbf{P}_\theta$.

Con otras palabras, en la clase K_b, la aplicación de la operación $\mathbf{M}_\theta(\cdot/S)$ a θ^* mejora uniformemente la estimación θ^*.

Demostración. El hecho de que θ_S^* es una estimación, significa que θ_S^* no depende de θ y que es una función medible de X. Su independencia respecto a θ se desprende de las propiedades de las características estadísticas, ya que la distribución de X para una S registrada no depende de θ ($\mathbf{M}_\theta(\theta^*/S)$, para la estadística arbitraria S, hablando en general, depende de θ). Al mismo tiempo, en virtud de las propiedades de la e.m.c., θ_S^* es una función medible de S y, por lo tanto, también de X. Por consiguiente, θ_S^* es la estimación que satisface la propiedad 2) del teorema.

La igualdad

$$\mathbf{M}_\theta\theta_S^* = \mathbf{M}_\theta\mathbf{M}_\theta(\theta^*/S) = \mathbf{M}_\theta\theta^*,$$

que demuestra que $\theta_S^* \in K_b$, también se deduce directamente de las propiedades de la e.m.c. Seguidamente,

$$\mathbf{M}_\theta(\theta^* - \theta)^2 = \mathbf{M}_\theta(\theta^* - \theta \pm \theta_S^*)^2 = \mathbf{M}_\theta(\theta_S^* - \theta)^2 + \mathbf{M}_\theta(\theta^* - \theta_S^*)^2 + \\ + 2\mathbf{M}_\theta(\theta_S^* - \theta)(\theta^* - \theta_S^*).$$

Utilizando de nuevo las propiedades de la e.m.c., obtenemos

$$\mathbf{M}_\theta(\theta_S^* - \theta)(\theta^* - \theta_S^*) = \mathbf{M}_\theta\mathbf{M}_\theta[(\theta_S^* - \theta)(\theta^* - \theta_S^*)/S] =$$
$$= \mathbf{M}_\theta[(\theta_S^* - \theta)\mathbf{M}_\theta(\theta^* - \theta_S^*/S)] = 0$$

y, por consiguiente,

$$\mathbf{M}_\theta(\theta^* - \theta)^2 - \mathbf{M}_\theta(\theta_S^* - \theta)^2 + \mathbf{M}_\theta(\theta^* - \theta_S^*)^2. \ \triangleleft$$

En realidad, la desigualdad 3) del teorema 1 se puede obtener directamente de la propiedad de la e.m.c., $(\mathbf{M}(\xi/S))^2 \leqslant \mathbf{M}(\xi^2/S)$, ya que entonces

$$(\theta_S^* - \theta)^2 = [\mathbf{M}_\theta(\theta^* - \theta)/S)]^2 \leqslant \mathbf{M}_\theta[(\theta^* - \theta)^2/S],$$
$$\mathbf{M}_\theta(\theta_S^* - \theta)^2 \leqslant \mathbf{M}_\theta(\theta^* - \theta)^2.$$

El hecho expuesto en el teorema 1 puede interpretarse del modo siguiente. Supongamos que S y T son dos estadísticas suficientes, $\theta^* = \varphi(T)$ y S está subordinada a T, entonces $\mathbf{M}_\theta(\theta_S^* - \theta)^2 \leqslant \mathbf{M}_\theta(\theta^* - \theta)^2$.

Con otras palabras, cuanto más "económica" sea la estadística suficiente S (o cuanto más pobre sea la σ-álgebra correspondiente), tanto mejores serán las estimaciones θ_S. Así pues, para construir las estimaciones óptimas debemos buscar las estadísticas suficientes *mínimas* (o las σ-álgebras mínimas). En este caso, en calidad de estimaciones iniciales θ^* también pueden figurar estimaciones "malas" que no poseen, por ejemplo, incluso propiedad de validez. En este sentido es aleccionador el siguiente

Ejemplo 1. Sea $X \in \Pi_\lambda$. La estimación $\lambda^* = x_1$, evidentemente, no está desplazada $\mathbf{M}\lambda^* = \mathbf{M}x_1 = \lambda$ ($b\lambda = 0$) y no es válida, ya que no depende de n. La estadística suficiente mínima de λ es la estadística $S = n\bar{x} = \sum x_i$. Del ejemplo 12.1 se deduce que la distribución x_1 condicional respecto a S es la distribución $\mathbf{B}_{1/n}^S$ en el esquema de Bernoulli, con una probabilidad de éxito igual a $1/n$:

$$\mathbf{P}(x_1 = k/S = s) = C_s^k\left(\frac{1}{n}\right)^k\left(1 - \frac{1}{n}\right)^{s-k}.$$

Por consiguiente,

$$\lambda_S^* = \mathbf{M}(x_1/S) = \sum_{k=1}^{S} kC_S^k\left(\frac{1}{n}\right)^k\left(1 - \frac{1}{n}\right)^{S-k} = \frac{S}{n} = \bar{x}.$$

En uno de los ejemplos ulteriores demostraremos que $\bar{x}$ es una estimación eficiente.

2. Caso multidimensional. Ahora obtendremos los análogos del teorema 1 para el caso multidimensional cuando θ y θ^* son vectores de R^k.

Al igual que en el caso unidimensional, el vector $b(\theta) = \mathbf{M}_\theta\theta^* - \theta$ será

el desplazamiento de la estimación θ^*, y por K_b designaremos la clase de todas las estimaciones con desplazamiento b.

Teorema 1A. *Sea S una estadística suficiente y $\theta^* \in K_b$. Entonces la estimación $\theta_S^* = \mathbf{M}_\theta(\theta^*/S)$ posee las propiedades*

1) $\theta_S^* \in K_b$.

2) *θ_S^* depende exclusivamente de $S(X)$,*

3) *la dispersión estándar de θ_S^* no supera la dispersión estándar de θ^* o bien, que es lo mismo, para cualquier vector $a \in R^k$*

$$\mathbf{M}_\theta(\theta_S^* - \theta, a)^2 \leqslant \mathbf{M}_\theta(\theta^* - \theta, a)^2. \tag{1}$$

Aquí, la igualdad (para todos los valores de a) es posible únicamente en el caso de $\theta^* = \theta_S^*$ c.d. *respecto a $\mathbf{P}_\theta$.*

Demostración. Las primeras dos afirmaciones son evidentes. Las desigualdades (1) se deducen del teorema 1, puesto que todo se reduce al examen de las estimaciones unidimensionales (θ^*, a) del parámetro (θ, a), y $\mathbf{M}_\theta[(\theta^*, a)/S] = (\theta_S^*, a)$. Si en (1), para todos los valores de a es válida esa igualdad, entonces, para cada a tendremos $(\theta_S^*, a) = (\theta^*, a)$ c.d. Esto precisamente significa que $\theta_S^* = \theta^*$ c.d. ◁

Ahora bien, en el caso multidimensional, las estadísticas suficientes desempeñan el mismo papel: la forma cuadrática $\sum \sigma_{ij} a_i a_j$, donde $\sigma^2 = |\sigma_{ij}|$ es la matriz de segundos momentos para $\theta_S^* - \theta$, será tanto menor cuanto menor sea la σ-álgebra de $\sigma(S)$ engendrada por S.

3. Estadísticas completas y estimaciones eficientes. Ahora citaremos un criterio muy simple del inmejoramiento de las estimaciones, basado en el concepto de plenitud de la característica S. Designemos por l la dimensión de la característica S. Esta suele ser mayor que la dimensión k del parámetro θ o igual a ésta.

Para dos funciones medibles $f_1(s)$ y $f_2(s): R^l \to R^k$ escribiremos $f_1(s) = f_2(s)$ c.d. $[\mathscr{P}]$, donde $\mathscr{P}$ es la familia de distribuciones en $(R^l, \mathfrak{B}^l)$ si $f_1(s) = f_2(s)$ en todas las partes excepto el conjunto N tal que $\mathbf{P}(N) = 0$ para todas $\mathbf{P} \in \mathscr{P}$.

Definición 2. La familia de distribuciones $\mathscr{G} = \{\mathbf{G}_\theta\}$ en $(R^l, \mathfrak{B}^l)$, que dependen del parámetro k-dimensional $\theta \in \Theta \subset R^k$, se llama *completa* si la igualdad

$$\int y(s) = \mathbf{G}_\theta(ds) = 0 \text{ cuando todos } \theta \in \Theta \tag{2}$$

conduce a $y(s) = 0$ c.d. $[\mathscr{G}]$. La ecuación (2) se examina en la clase de funciones y: $R^l \to R^k$ para las cuales existe la integral (2).

Definición 3. La estadística S se denomina *completa* si la familia $\mathscr{G}$ de sus distribuciones $\mathbf{G}_\theta$, inducidas por la distribución $\mathbf{P}_\theta$ en $(\mathscr{X}^n, \mathfrak{B}^n_{\mathscr{X}})$, es completa.

La ecuación (2) para las estadísticas puede ser escrita en forma de $\mathbf{M}_\theta y(S) = 0$ para todos $\theta \in \Theta \subset R^k$.

Teorema 2. *La estadística S es completa si y sólo si para cualquier $b_0(\theta)$, la $\sigma(S)$-medible* [*)] *estimación θ^* es única en la clase de todas las $\sigma(S)$-medibles estimaciones de K_{b_0}.*

Si la $\sigma(S)$-medible estimación es única en K_{b_0}, entonces las $\sigma(S)$-medibles estimaciones también poseerán la propiedad de unicidad en cualquier otra clase K_b.

La *demostración* de esta afirmación es casi evidente, ya que la existencia de dos $\sigma(S)$-medibles estimaciones $\theta_1^* = \varphi_1(S)$ y $\theta_2^* = \varphi_2(S)$ en K_{b_0} significa que $\int \varphi_i(s)\mathbf{G}_\theta(ds) = b_0(\theta)$, $i = 1, 2$,

$$\int [\varphi_1(s) - \varphi_2(s)]\mathbf{G}_\theta(ds) = 0 \text{ para todos } \theta \in \Theta,$$

así que la plenitud de S conduce a $\varphi_1(s) = \varphi_2(s)$ c.d. $[\mathcal{G}]$. Al contrario, sea $\int y(s)\mathbf{G}_\theta(ds) = 0$ para todos $\theta \in \Theta$, $\theta_1^* = \varphi_1(s) \in K_b$. Entonces $\theta_2^* = \varphi_1(s) + y(s) \in K_b$, y la unicidad de la $\sigma(S)$-medible estimación significa que $y(s) = 0$ c.d. $[\mathcal{G}]$. ◁

Teorema 3. *Si la estadística suficiente S es completa, y $\theta^* \in K_b$, entonces la estimación $\theta_S^* = \mathbf{M}_\theta(\theta^*/S)$ es la estimación eficiente única en K_b.*

Este teorema nos ofrece criterios suficientemente simples de eficacia de las estimaciones.

Demostración. En virtud del teorema (2), la $\sigma(S)$-medible estimación en la clase K_b es única.

Sea θ^{**} cualquier otra estimación de K_b. Entonces $\theta_S^{**} = \mathbf{M}_\theta(\theta^{**}/S) \in K_b$ y, por lo tanto, $\Theta_S^* = \theta_S^*$ c.d. $[\mathcal{G}]$. De aquí y del teorema 1 se desprende que

$$\mathbf{M}_\theta(\theta_S^* - \theta)^2 = \mathbf{M}_\theta(\theta_S^{**} - \theta)^2 \leqslant \mathbf{M}_\theta(\theta^{**} - \theta)^2,$$

y la igualdad es posible únicamente para $\theta^{**} = \theta_S^*$ c.s. ◁

Corolario 1. *Si S es una estadística suficiente completa, y θ^* es una estimación no desplazada, entonces θ_S^* es una estimación eficiente y es la única.*

Ejemplo 2. En el ejemplo 1, con distribución de Poisson, hemos obtenido que para $\lambda^* = x_1$

$$\lambda_S^* = \mathbf{M}_\lambda(x_1/S) = \bar{x},$$

donde $S = n\bar{x}$. Mostremos que S es una estadística completa y, por consiguiente, $\bar{x}$ es una estimación suficiente. La ecuación (2) para la estadística

---

[*)] O sea, medible respecto a la σ-álgebra de $\sigma(S)$ engendrada por S y, por lo tanto, representable en forma de $\varphi(S)$, donde φ es la función de Borel.

S tiene la forma

$$\sum_{k=0}^{\infty} y(k) e^{-n\lambda} \frac{(n\lambda)^k}{k!} = 0 \text{ cuando todos } \lambda \geqslant 0,$$

o, que es lo mismo,

$$v(z) = \sum y(k) \frac{z^k}{k!} = 0 \text{ para todos } z \geqslant 0. \quad (3)$$

Es evidente que esto conduce a $y(k) = 0$, ya que de la convergencia de la serie (3), digamos, cuando $z = 1$ se deduce que $v(z)$ es analítica cuando $|z| < 1$ y es idénticamente igual a 0. Por consiguiente, los coeficientes $y(k)$ de su desarrollo en serie son iguales a 0.

Ejemplo 3. Sea $X \in \mathbf{U}_{0,\theta}$. Mostremos que la estadística $S = x_{(n)} = \max_{i \leqslant n} x_i$ es completa. La suficiencia (y minimización) de S ha sido establecida en el ejemplo 13.2. La distribución de S se define por la igualdad

$$\mathbf{P}(S < s) = (s/\theta)^n, \quad 0 \leqslant s \leqslant \theta,$$

así que S tiene una densidad igual a $ns^{n-1}\theta^{-n}$ cuando $s \in [0, \theta]$. En este caso la ecuación (2) tiene la forma

$$\int_0^\theta y(s) \frac{ns^{n-1}}{\theta^n} ds = 0 \text{ cuando } \theta \in (0, \infty).$$

De la igualdad $\int_0^\theta y(s)s^{n-1}ds = 0$ para todos θ resulta, evidentemente, que $y(s)s^{n-1} = 0$, $y(s) = 0$ c.d.

Le proponemos al lector que verifique si son completas las estadísticas suficientes para otras familias paramétricas y, en particular, que determine si $\alpha^* = \frac{1}{\bar{x}}\left(1 - \frac{1}{n}\right)$ es la estimación eficiente única del parámetro α de la familia $\Gamma_{\alpha,1}$ (véase § 2).

Señalemos ahora que el teorema 3 muestra la existencia de relaciones entre los conceptos de amplitud y minimización. En este aspecto es válida la afirmación siguiente, que da, junto con los teoremas del § 13, el criterio de minimización de las estadísticas suficientes.

Teorema 4. *Cualquier característica suficiente completa S es una estadística suficiente mínima.*

Demostración. Sea $\mathfrak{U}_0$ una σ-álgebra suficiente mínima (según el teorema 13.1, ésta existe). Supongamos que $\mathbf{M}_\theta S$ existe y examinemos la función

$\psi = S - \mathbf{M}_\theta(S/\mathfrak{U}_0)$. Como $\mathfrak{U}_0 \subset \sigma(S)$, entonces ψ será $\sigma(S)$-medible, así que $\psi = \psi(S)$. Designemos por $\mathbf{G}_\theta$ la distribución de S. Entonces es evidente que para todos θ, $\mathbf{M}_\theta\psi(S) = 0$ o, que es lo mismo,

$$\int \psi(s)\mathbf{G}_\theta(ds) = 0 \text{ para todos } \theta \in \Theta.$$

De aquí, en virtud de la amplitud de S resulta que $\psi(s) = 0$ c.s. $[\mathscr{G}]$, $\mathscr{G} = \{\mathbf{G}_\theta\}$. Esto significa que $S = \mathbf{M}_\theta(S/\mathfrak{U}_0)$ c.d. $[\mathscr{G}]$ y, por lo tanto, S es medible respecto a*) $\mathfrak{U}_0$, $\sigma(S) = \mathfrak{U}_0$.

Si $\mathbf{M}_\theta S$ no existe, es necesario, en vez de S, examinar la estadística arctg S, la cual es, evidentemente, equivalente a S en cuanto a las propiedades de suficiencia, amplitud y minimización. ◁

Señalemos que la afirmación inversa no es cierta: la *estadística suficiente mínima no es obligatoriamente completa*. Los ejemplos respectivos se obtienen fácilmente en los casos en que la dimensión l de la estadística es mayor que la dimensión k del parámetro θ. Por ejemplo, en el § 13 hemos visto que la densidad compatible de la estadística suficiente mínima $S = (x_{(1)}, x_{(n)})$ para la familia $\mathbf{U}_{\theta, 1+\theta}$ es igual a

$$g_\theta(u, v) = \begin{cases} n(n-1)(v-u)^{n-2} & \text{cuando } u \geqslant \theta,\ v \leqslant 1+\theta,\ v > u, \\ 0 & \text{en los demás casos.} \end{cases}$$

Si se toma la función $y(u, v) = \varphi(v - u)$ y se hace la transformación ortogonal $(v - u)/\sqrt{2} = t$, $(v + u)/\sqrt{2} = z$, la integral en (2) por el triángulo $u \geqslant \theta$, $v \leqslant 1 + \theta$, $v > u$) será igual a

$$\int y(u, v) g_\theta(u, v)\,du\,dv = n(n-1)\int_0^1 \varphi(x)x^{n-2}(1-x)\,dx.$$

Es evidente que la integral en el segundo miembro no depende de θ y es fácil elegir la función $\varphi(x) \not\equiv 0$ que la reduce a cero.

§ 15. Familia exponencial

Supongamos que $\theta = (\theta_1, \ldots, \theta_k)$ es un parámetro k-dimensional y que la densidad $f_\theta(x)$ es representable en la forma

$$f_\theta(x) = h(x) \exp\left\{\sum_{j=1}^{k} a_j(\theta)U_j(x) + V(\theta)\right\}, \tag{1}$$

donde todas las funciones que entran en el segundo miembro son finitas y medibles.

---

*) Por $\mathfrak{U}_0$ aquí es necesario entender la σ-álgebra completada por los conjuntos N, para los cuales $\mathbf{P}_\theta(N) = 0$ para todos θ.

Definición 1. Las familias de distribuciones $\{\mathbf{P}_\theta\}$, con densidad de este género, se llaman *familias exponenciales* y se designan con el símbolo $\mathscr{E}$.

Para hacer que la representación (1) sea, en la medida de lo posible, unívoca, supondremos que las *funciones* $a_0(\theta) = 1$, $a_1(\theta)$, ..., $a_k(\theta)$ *son linealmente independientes en* Θ.

Como veremos, las familias exponenciales ocupan un lugar especial entre las familias paramétricas de distribuciones, ya que para ellas muchas construcciones generales de la estadística matemática pueden ser realizadas en forma explícita.

A veces se llaman familias exponenciales las familias de distribuciones de tipo más particular *), cuando $a_j(\theta) = \theta_j$.

A las familias exponenciales pertenecen, por ejemplo, las familias de distribuciones $\{\Phi_{\alpha,\sigma^2}\}$, $\{\Pi_\lambda\}\{\mathbf{B}_p\}$, $\{\Gamma_{\alpha,\lambda}\}$ y una serie de otras.

Ejemplo 1. Examinemos la distribución $\Gamma_{\alpha,\lambda}$. Su densidad $\gamma_{\alpha,\lambda}(x)$ se puede representar en la forma

$$\gamma_{\alpha,\lambda}(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1}e^{-\alpha x} = x^{-1}\exp\left\{\lambda \ln x - \alpha x + \ln \frac{\alpha^\lambda}{\Gamma(\lambda)}\right\},\ x > 0,$$

así que aquí se puede poner

$$h(x) = \begin{cases} x^{-1}, & x > 0, \\ 0, & x \leqslant 0, \end{cases}$$

$$U_1(x) = \ln x,\ U_2(x) = x,\ V(\alpha, \lambda) = \ln \frac{\alpha^\lambda}{\Gamma(\lambda)},$$

$$a_1(\alpha, \lambda) = \lambda,\ a_2(\alpha, \lambda) = -\alpha. \ \triangleleft$$

La función de verosimilitud para $X \in \mathbf{P} \in \mathscr{E}$ es igual a

$$f_\theta(X) = \exp\{(a(\theta), S) + nV(\theta)\}\prod_{i=1}^n h(x_i),$$

donde

$$a(\theta) = (a_1(\theta), \ldots, a_k(\theta)),\ S = (S_1, \ldots, S_k),$$

$$S_j = S_j(X) = \sum_{i=1}^n U_j(x_i),$$

(a, S) es el producto escalar. De aquí y del teorema 12.1 resulta que S es una función suficiente para θ. Demostremos que S es una estadística suficiente mínima.

---

*) En realidad, esto es lo mismo; llegaremos a una forma particular si realizamos la transformación biunívoca $\gamma = \gamma(\theta)$, $\gamma = \gamma_1, \ldots, \gamma_k)$ sobre el parámetro θ, poniendo $\gamma_j = a_j(\theta)$.

Como las funciones $a_j(\theta)$, $U_j(x)$, $V(\theta)$ son finitas, la exponencial en (1) es siempre positiva. Esto significa que en calidad de distribución **Q** en el teorema 13.1 (con la que todas las $\mathbf{P}_\theta$ son absolutamente continuas respecto a $\mathbf{P}_Q = \int \mathbf{P}_t \mathbf{Q}(dt)$) se puede tomar la distribución concentrada en cualquier punto fijado θ^0. Por eso, del teorema 13.1 se deduce que la σ-álgebra de $\mathfrak{U}_0$ engendrada por la función

$$r(X, \theta) = \frac{f_\theta(X)}{f_{\theta^0}(X)} = \exp\{(a(\theta) - a(\theta^0), S) + n(V(\theta) - V(\theta^0))\}$$

es la σ-álgebra suficiente mínima.

Teorema 1. *La estadística S es una estadística suficiente mínima.*

Demostración. De la independencia lineal de las funciones $1, a_1(\theta), \ldots, a_k(\theta)$ en Θ se deduce la independencia lineal $a_1(\theta) - a_1(\theta^0), \ldots, a_k(\theta) - a_k(\theta^0)$. Esto significa que en Θ hay k puntos $\theta^1, \ldots, \theta^k$ tales que los valores $a_{ij} = a_i(\theta^j) - a_i(\theta^0)$ forman una matriz A cuya determinante se distingue del cero. Esto significa, a su vez, que las ecuaciones $(a(\theta^j) - a(\theta^0), S) = \ln r(X, \theta^j) - n(V(\theta^j) - V(\theta^0))$, $j = 1, \ldots, k$, son solubles unívocamente respecto a S y, por lo tanto, $\sigma(S) \subset \sigma(r(X, \theta_j); j = 1, \ldots, k) \subset \mathfrak{U}_0$. $\triangleleft$

En el ejemplo 1 hemos examinado la distribución Γ y establecimos que para ésta es válida la representación (1) cuando $\theta = (\alpha, \lambda)$ con las funciones

$$U_1(x) = \ln x, \quad U_2(x) = x,$$
$$a_1(\alpha, \lambda) = \lambda, \quad a_2(\alpha, \lambda) = -\alpha.$$

Es evidente que las condiciones del teorema 1 se han cumplido y que la estadística $S = \left(\sum \ln x_i, \sum x_i\right)$ o bien, que es lo mismo, la estadística $(\Pi x_i, \Sigma x_i)$ es una estadística suficiente mínima.

Si reforzamos un poco las condiciones del teorema 1, entonces la estadística S será una estadística suficiente completa (en este caso la minimización de S se podría obtener como consecuencia de la plenitud).

Teorema 2. *Sea $X \in \mathbf{P} \in \mathscr{E}$. Si la función a y el conjunto Θ son tales que $a(\theta)$ traza un paralelepípedo k-dimensional cuando θ recorre Θ, entonces S es una estadística suficiente completa.*

Es evidente que las condiciones del teorema respecto al paralelepípedo se cumplirán si el conjunto Θ es "sólido", es decir, si contiene los puntos interiores (y junto con ellos también las esferas en R^k, de radio bastante pequeño) y si en el entorno de cualquier punto "sólido" θ^0, las funciones $a_j(\theta)$ son linealmente independientes y suaves. Entonces la transformación $a = a(\theta)$ transfiere el entorno del punto θ^0 al conjunto sólido.

Es evidente que el ejemplo 1, con la distribución Γ, satisface las condiciones del teorema 2, ya que la estadística $(\Pi x_i, \Sigma x_i)$ es completa.

De un modo igualmente sencillo, el lector puede comprobar que para la distribución normal Φ_{α,σ^2}, la estadística $\left(\sum x_i, \sum x_i^2\right)$ también es una estadística suficiente completa.

Demostración del teorema 2. En nuestro caso las funciones $\psi(s, \theta)$ y $h(x)$ en el teorema de factorización de Neyman — Fisher son iguales a

$$\psi(s, \theta) = \exp\{ (a(\theta), s) + nV(\theta) \},$$

$$h(x) = \prod_{i=1}^{n} h(x_i).$$

Examinemos en $(R^k, \mathfrak{B}^k)$ la medida que no depende de θ:

$$\nu(B) = \int_{S^{-1}(B)} h(x)\mu^n(dx),$$

donde $S^{-1}(B)$ es el conjunto de todos los x para los cuales $S(x) \in B$.

Destaquemos en forma de lemas, las dos siguientes afirmaciones auxiliares.

Lema 1. *La distribución* $\mathbf{G}_\theta(B) = \mathbf{P}_\theta(S(X) \in B)$ *de la estadística* S *es absolutamente continua respecto a* ν, *y en el punto* s *tiene una densidad igual a* $\psi(s, \theta)$.

La **demostración** se deduce de la igualdad

$$\mathbf{G}_\theta(B) = \int_{S(x)\in B} \psi(S(x), \theta) h(x)\mu^n(dx) = \int_{s\in B} \psi(s, \theta)\nu(ds),$$

la cual es consecuencia de la sustitución de las variables. ◅

Lema 2. *Sean* $\mathbf{G}_1$ *y* $\mathbf{G}_2$ *dos medidas* σ*-finitas en* $(R^k, \mathfrak{B}^k)$. *En este caso, si* $\int e^{(a,u)}\mathbf{G}_1(du) = \int e^{(a,u)}G_2(du)$ *existen para todos los valores de* a *de cierto paralelepípedo* I *en* R^k, *entonces* $\mathbf{G}_1 = \mathbf{G}_2$.

Demostración. Para simplificar los razonamientos examinemos el caso unidimensional $k = 1$ y supongamos que $I = \{x: |x| \leqslant \alpha\}$. Entonces

$$h_j(a) = \int e^{au}\mathbf{G}_j(du), \; j = 1, 2,$$

son funciones analíticas cuando $|a| < \alpha$. Además, para todos $b \in R$ están definidas las funciones $h_j(z) = \int e^{(a+ib)u}\mathbf{G}_j(du)$ de la variable compleja $z = a + ib$. Naturalmente que $h_j(z)$ serán analíticas en la franja de $|a| < \alpha$, $-\infty < b < \infty$. Como $h_1(z) = h_2(z)$ en el segmento de la recta $b = 0$, $|a| < \alpha$, entonces $h_1(z) = h_2(z)$ para todas z de la franja indicada. Por lo tanto,

$$\int e^{ibu}\mathbf{G}_1(du) = \int e^{ibu}\mathbf{G}_2(du). \quad (2)$$

Señalemos que en vista de que $h_j(0) = \int G_j(du) < \infty$, podemos considerar que G_j son medidas probabilísticas. Del teorema de la correspondencia biunívoca entre las funciones características y las distribuciones [11], así como de (2), resulta que $G_1 = G_2$.

Si el paralelepípedo I tiene la forma $\{x: |x - \alpha_0| \leqslant \alpha\}$, entonces conviene pasar a las medidas $G_j^*(du) = e^{\alpha_0 u} G_j(du)$.

En el caso multidimensional $k > 1$, la demostración se realiza exactamente igual. ◁

Ahora podemos pasar directamente *a la demostración del teorema 2.*

Debemos demostrar que si φ es una función medible en $(R^k, \mathfrak{B}^k)$ y existe

$$\int \varphi(s) G_\theta(ds) = 0 \text{ para todos } \theta \in \Theta, \tag{3}$$

entonces $\varphi(s) = 0$ c.d. $[\mathscr{G}]$, $\mathscr{G} = \{G_\theta\}_{\theta \in \Theta}$. Sea $\varphi = \varphi^+ - \varphi^-$, donde $\varphi^\pm \geqslant 0$. En este caso, de (3) se desprende $\int \varphi^+(s) G_\theta(ds) = \int \varphi^-(s) G_\theta(ds)$ o bien, en virtud del lema 1,

$$\int \varphi^+(s) \psi(s, \theta) \nu(ds) = \int \varphi^-(s) \psi(s, \theta) \nu(ds),$$
$$\int \varphi^+(s) e^{(s, a(\theta))} \nu(ds = \int \varphi^-(s) e^{(s, a(\theta))} \nu(ds).$$

Si formamos las medidas σ-finitas $\nu^\pm(ds) = \varphi^\pm(s)\nu(ds)$, obtendremos

$$\int e^{(s, a)} \nu^+(ds) = \int e^{(s, a)} \nu^-(ds)$$

para todos los valores de a de cierto paralelepípedo en R^k. Sólo nos queda hacer uso del lema 2. ◁

Corolario 1. *Si* $X \Subset P \in \mathscr{E}$, $\theta^* \in K_b$ *y se cumplen las condiciones del teorema 2, la estimación* $\theta_S^* = M(\theta^*/S)$ *es la estimación eficiente en* K_b.

§ 16. Desigualdad de Rao — Cramer y estimaciones R-eficientes

1. Desigualdad de Rao — Cramer y sus corolarios. Los resultados de los párrafos precedentes nos proporcionaron varios criterios de eficacia de las estimaciones. Sin embargo, estos criterios tenían, en cierto sentido, un carácter cualitativo. En este párrafo continuaremos el estudio de la cuestión acerca de las estimaciones eficientes, pero desde un punto de vista un poco diferente. Aclaremos, ante todo, cuál es el valor mínimo del error estándar que se puede obtener.

Al principio examinaremos el caso unidimensional cuando θ es un parámetro escalar. Con respecto al conjunto Θ, para precisar vamos a suponer que eso es un intervalo finito o infinito, cerrado o abierto.

Para responder a la pregunta planteada necesitaremos las condiciones de regularidad en $f_\theta(x)$. Sea, como antes,

$$l(x, \theta) = \ln f_\theta(x), \ L(X, \theta) = \sum_{i=1}^{n} l(x_i, \theta), \ a(\theta) = \mathbf{M}_\theta \theta^* = \theta + b(\theta).$$

Supongamos que se ha cumplido la condición (R). *Las funciones* $\sqrt{f_\theta(x)}$ *para c.t.*[μ] *valores de x son continuamente derivables respecto a* $\theta \in \Theta$, *y la integral*

$$I(\theta) = \int \frac{(f_\theta'(x))^2}{f_\theta(x)} \mu(dx) = \mathbf{M}_\theta [l'(x_1, \theta)]^2 \qquad (1)$$

existe y es positiva y continua según θ. (Aquí y en lo sucesivo, la tilde significa la derivación respecto a θ).

Con arreglo a la integral (1) es necesario señalar lo siguiente: Si x, junto con su entorno, no pertenece al portador $N_{P_\theta} = \{x; f_\theta(x) > 0\}$ de la distribución $\mathbf{P}_\theta$, entonces la función subintegral $(f_\theta'(x))^2/f_\theta(x)$ se convierte en indeterminación de tipo 0/0. Convendremos en considerar esta razón igual a cero. Seguiremos esa misma regla en cuanto a la derivada $l'(x, \theta) = f_\theta'(x)/f_\theta(x)$, al integrarla. Podríamos no hacer estas restricciones si desde el principio eximinaramos las integrales de la forma de $\mathbf{M}_\theta \varphi(x_1, \theta)$ sólo en la región de N_{P_θ}.

La función $I(\theta)$ es conocida con el nombre de *información de Fisher* y desempeña un papel muy importante en la matemática estadística, además, en lo sucesivo tropezaremos repetidas veces con ella. Algunas propiedades de la función $I(\theta)$ se examinan en § 17.

Si el conjunto Θ es compacto, la *continuidad de* $I(\theta)$ *en las condiciones* (R) *es equivalente a la condición*

$$\sup_{\theta \in \Theta} \mathbf{M}_\theta([l'(x_1, \theta)]^2; \ |l'(x_1, \theta)| > N) \to 0$$

cuando $N \to \infty$, la cual se puede llamar convergencia uniforme de la integral $I(\theta)$ (véase el Suplemento VI).

Tiene lugar la siguiente desigualdad para la varianza de las estimaciones θ^* con desplazamiento b.

Teorema 1 (desigualdad de Rao — Cramer). *Si* $\theta^* \in K_b$ *y si está cumplida la condición* (R) *y* $\mathbf{M}_\theta(\theta^*)^2 < c < \infty$, *entonces*

$$\mathbf{D}_\theta \theta^* \geqslant \frac{[1 + b'(\theta)]^2}{nI(\theta)}. \qquad (2)$$

Si en dicha desigualdad se alcanza igualdad en cierto segmento $\theta \in [\theta_1, \theta_2] \subset \Theta$, *y* $\mathbf{D}_\theta \theta^* > 0$ *en ese segmento, entonces la función de verosimilitud*

$f_\theta(X)$ para $\theta \in [\theta_1, \theta_2]$ es representable en la forma

$$f_\theta(X) = \exp\{\theta^* A(\theta) + B(\theta)\} h(X), \tag{3}$$

donde $A(\theta)$, $B(\theta)$ no dependen de X.

Al contrario, si θ^* = const, o si es válida la representación (3), *entonces en la desigualdad* (2) *se alcanza igualdad.*

Evidentemente, la condición (3) significa que la distribución en $\mathscr{X}^n$ con densidad $f_\theta(x)$ pertenece a la familia exponencial $\mathscr{E}$.

Corolario 1. *Si se cumplen las condiciones del teorema* 1,

$$\mathbf{M}_\theta(\theta^* - \theta)^2 \geqslant \frac{[1 + b'(\theta)]^2}{nI(\theta)} + b^2(\theta).$$

Para cualquier estimación no desplazada θ^*,

$$\mathbf{M}_\theta(\theta^* - \theta)^2 \geqslant \frac{1}{nI(\theta)}.$$

Así pues, en las clases K_b, el valor mínimo posible de las desviaciones estándar es distinto de cero y se define por los segundos miembros de las desigualdades escritas.

Observación 1. En cuanto a la condición $\mathbf{M}_\theta(\theta^*)^2 < c < \infty$ se puede notar que cuando $\mathbf{M}_\theta(\theta^*)^2 = \infty$ se cumple $\mathbf{D}_\theta\theta = \infty$ y la desigualdad (2) se vuelve trivial. En virtud de (2), la condición $\mathbf{D}_\theta\theta > 0$ se puede sustituir por $(1 + b'(\theta))^2 > 0$.

Observación 2. A la par con la condición (R) se pueden señalar algunas otras condiciones que aseguran la afirmación del teorema 1 y que se distinguen muy poco una de otra. Nos hemos detenido en aquellas de ellas que nos serán más cómodas en los párrafos posteriores. Las condiciones de tipo algo diferente se citarán en el § 22.

Necesitaremos una afirmación auxiliar.

Lema 1. *Supongamos que se ha cumplido la condición* (R) y *que $S = S(X)$ es cualquier estadística para la cual $\mathbf{M}_\theta S^2 < c < \infty$ cuando $\theta \in \Theta$. Entonces la función*

$$a_S(\theta) = \mathbf{M}_\theta S = \int S(x) f_\theta(x) \mu^n(dx) \tag{4}$$

es derivable respecto a θ, además

$$a_S'(\theta) = \int S(x) f_\theta'(x) \mu^n(dx) = \mathbf{M}_\theta S L'(X, \theta). \tag{5}$$

Esta afirmación tiene carácter técnico y su demostración dificultaría considerablemente las investigaciones. Por eso hemos pasado la demostración del lema 1 al Suplemento VI.

Demostración del teorema 1. Poniendo en (5) $S \equiv 1$, obtenemos $a_S(\theta) \equiv 1$,

$$\mathbf{M}_\theta L' = 0, \ \mathbf{M}_\theta a(\theta) L' = 0. \tag{6}$$

Volviendo a utilizar (5) para $S = \theta^*$ y (6), obtenemos

$$\mathbf{M}_\theta \theta^* L' = a'(\theta), \ \mathbf{M}_\theta(\theta^* - a(\theta))L' = a'(\theta). \tag{7}$$

Según la desigualdad de Cauchy — Buniakovski,

$$(a'(\theta))^2 \leqslant \mathbf{M}_\theta(\theta^* - a(\theta))^2 \mathbf{M}_\theta(L')^2 \tag{8}$$

o bien, que es lo mismo,

$$\mathbf{D}_\theta \theta^* \geqslant \frac{(1 + b'(\theta))^2}{\mathbf{M}_\theta(L')^2}. \tag{9}$$

Como las variables aleatorias $l_j = l'(\mathbf{x}_j, \theta)$ son independientes, están igualmente distribuidas y tienen, en virtud de (6), una esperanza matemática nula, $\mathbf{M}_\theta l_j = 0$, entonces $\mathbf{M}_\theta l_i l_j = 0$ cuando $i \neq j$,

$$\mathbf{M}_\theta(L')^2 = \mathbf{M}_\theta\left(\sum_j l_j\right)^2 = \sum_{i,j} \mathbf{M}_\theta l_i l_j = n\mathbf{M}_\theta l_1^2 = nI(\theta).$$

Junto con (9) esto demuestra la desigualdad (2).

Demostremos ahora la segunda afirmación del teorema. Para simplificar la demostración consideraremos que Θ coincide con $[\theta_1, \theta_2]$ y que la medida μ está concentrada en la unión de los portadores de $\mathbf{P}_\theta$, $\theta \in \Theta$. El signo de igualdad en (2) (o en 8)) quiere decir que

$$\int (\theta^* - a(\theta)) f'_\theta(x) \mu^n(dx) = \left[\int (\theta^* - a(\theta))^2 f_\theta(x) \mu^n(dx) \int \frac{(f'_\theta(x))^2}{f_\theta(x)} \mu^n(dx)\right]^{1/2}$$

para todos $\theta \in \Theta$. En vista de que la primera integral en el segundo miembro es positiva, la igualdad escrita sólo será posible si

$$f'_\theta(x)/\sqrt{f_\theta(x)} = c(\theta)(\theta^* - a(\theta))\sqrt{f_\theta(x)} \ \text{c.t.}[\mu^n]. \tag{10}$$

Designemos por A el conjunto de x para los que está cumplida (10) y $|\theta^*| < \infty$. Entonces $\mu(\overline{A}) = 0$ ($\overline{A}$ es el complemento a A). Anotamos $x \in A$. En virtud de la continuidad $f_\theta(x)$ en θ, tendremos $f_1(x) > 0$ en cierto intervalo $(t_1, t_2) \subset \Theta$, y en este intervalo, en virtud de (10),

$$L'(x, \theta) = c(\theta)(\theta^* - a(\theta)). \tag{11}$$

Señalemos ahora, que de (7), (11) y (2) resulta

$$a'(\theta) = \mathbf{M}_\theta(\theta^* - a(\theta))L' = c(\theta)\mathbf{D}_\theta\theta^*, \ \mathbf{D}_\theta\theta^* = \frac{(a'(\theta))^2}{nI(\theta)}, \tag{12}$$

$$|c(\theta)| = \sqrt{\frac{nI(\theta)}{\mathbf{D}_\theta \theta^*}},$$

así que $\mathbf{D}_\theta \theta^*$ es continua en θ junto con $a'(\theta)$, $I(\theta)$, y $|c(\theta)|$ junto con $a(\theta)$ están limitadas uniformemente en $[\theta_1, \theta_2]$. La derivada $L'(x, \theta)$ en (11) posee esa misma propiedad. Pero esto significa que $L(x, t)$ es finita y que $f_\theta(x) > 0$ en todas las partes de $\Theta = [\theta_1, \theta_2]$, así que (11) es válida para todos θ. Integrando (11) dentro de los límites de θ_1 y θ, obtendremos

$$L(x, \theta) = \theta^* \int_{\theta_1}^{\theta} c(t)dt - \int_{\theta_1}^{\theta} c(t)a(t)dt + L(x, \theta_1),$$

que es equivalente a (3) para $[\mu^n]$ c.t. x. Como la variación $f_\theta(x)$ en el conjunto de la μ^n-medida 0 no tiene importancia, (3) queda demostrada.

Examinemos ahora la última afirmación del teorema. Si $\theta^* =$ const, entonces $b'(\theta) = -1$ y ambos miembros de la desigualdad (2) se anulan. Supongamos que se ha cumplido (3). Entonces, derivando la función $L(X, \theta)$ respecto a θ, obtendremos

$$L'(X, \theta) = \theta^* A'(\theta) + B'(\theta).$$

De (7) se deduce que $a(\theta)A'(\theta) + B'(\theta) = 0$. Por eso

$$L'(X, \theta) = A'(\theta)(\theta^* - a(\theta))$$

y, por consiguiente (véase (10)), en (2) se alcanza la igualdad. ◁

En lo sucesivo excluiremos de las investigaciones el caso trivial $\theta^* =$ const y supondremos que $\mathbf{D}_\theta \theta^* > 0$ en todas las partes de Θ. Entonces es válido el

Corolario 2. *Si se cumplen las condiciones (R), para alcanzar la frontera inferior en la desigualdad de Rao — Cramer es necesario y suficiente que la estimación θ^* sea suficiente y que la función $\psi(\theta^*, \theta)$ en la igualdad de factorización tenga la forma*

$$\psi(\theta^*, \theta) = \exp\{\theta^* A(\theta) + B(\theta)\},$$

donde $A(\theta)$ y $B(\theta)$ son funciones derivables.

Corolario 3. *Si se cumplen las condiciones (R), $\theta^* \in K_b$, y en la desigualdad de Rao — Cramer se alcanza igualdad, entonces θ^* es una estimación eficiente en K_b.*

Esta afirmación se deduce de la representación

$$\mathbf{M}_\theta(\theta^* - \theta)^2 = \mathbf{D}_\theta \theta^* + b^2(\theta).$$

Señalemos que, hablando en general, lo contrario no es cierto: la estimación

puede ser eficiente en K_b, pero la frontera inferior $\frac{(1 + b'(\theta))^2}{nI(\theta)}$ para la varianza puede no alcanzarse.

Ejemplo 1. Sea $X \in \Gamma_{\alpha,1}$. Aquí $f_\alpha(X) = \alpha^n e^{-\alpha n \bar{x}}$. Las condiciones (R) en la región $\Theta \subseteq \{\alpha \geqslant \delta > 0\}$ están cumplidas. Es evidente que $S = n\bar{x}$ es una estadística suficiente completa. Por eso la estimación $\alpha^* = \bar{x}^{-1} = \mathbf{M}_\alpha(\bar{x}^{-1}/S)$ es eficiente en la clase K_b con un desplazamiento $b(\alpha) = \mathbf{M}_\alpha \bar{x}^{-1} - \alpha$.

Notemos ahora que $S \in \Gamma_{\alpha,n}$, así que cuando $n > 1$ (véase el § 2), $\mathbf{M}_\alpha \bar{x}^{-1} = n\mathbf{M}_\alpha S^{-1} = \frac{n}{n-1}\alpha$.

Ahora bien, la estimación $\alpha^{**} = \frac{n-1}{n\bar{x}} = \alpha^*\left(1 - \frac{1}{n}\right)$ no estará desplazada cuando $n > 1$. Análogamente, cuando $n > 2$ hallamos (véase el § 2 y también ejemplo 4.1)

$$\mathbf{M}_\alpha(\alpha^{**})^2 = (n-1)^2 \mathbf{M}_\alpha S^{-2} = \frac{n-1}{n-2}\alpha^2,$$

$$\mathbf{D}_\alpha \alpha^{**} = \alpha^2\left[\frac{n-1}{n-2} - 1\right] = \frac{\alpha^2}{n-2}.$$

Así pues, cuando $n > 2$, la estimación α^{**} es eficiente. Sin embargo, el criterio (3) no se ha cumplido, ya que

$$f_\alpha(X) = \alpha^{-n} e^{-\alpha(n-1)/\alpha^{**}}.$$

Por consiguiente, en la desigualdad de Rao — Cramer no se alcanza la frontera inferior. De esto también podemos convencernos directamente. En efecto, aquí $l(x, \alpha) = \ln \alpha - \alpha x$, $l'(x, \alpha) = 1/\alpha - x$ e

$$I(\alpha) = \mathbf{M}_\alpha[l'(x_1, \alpha)]^2 = \mathbf{M}_\alpha\left(\frac{1}{\alpha} - x_1\right)^2 = \frac{1}{\alpha^2} - \frac{2}{\alpha^2} + \frac{2}{\alpha^2} = \frac{1}{\alpha^2}.$$

Por lo tanto, cuando $n > 2$,

$$\frac{1}{nI(\theta)} = \frac{\alpha^2}{n} < \frac{\alpha^2}{n-2} = \mathbf{D}_\alpha \alpha^{**}.$$

Ahora bien, el logro de la frontera inferior en (2) es una exigencia más severa que el logro de la eficacia.

2. Estimaciones *R*-eficientes y asintóticamente *R*-eficientes. Supongamos que se han cumplido las condiciones (R). En este caso, el logro de la frontera inferior (exacto o asintótico) para la varianza en la desigualdad de Rao — Cramer puede ser un índice muy importante de la calidad de las estimaciones, íntimamente ligado al concepto de eficacia.

Definición 1. La estimación $\theta^* \in K_b$, para la cual

$$\mathbf{M}_\theta(\theta^* - \theta)^2 = \frac{(1 + b'(\theta))^2}{nI(\theta)} + b^2(\theta),$$

se llama *R-eficiente* (o *regularmente eficiente*) *en la clase* K_b.

La estimación R-eficiente en la clase K_0 de las estimaciones no desplazadas se denomina simplemente *R-eficiente*.

La estimación θ^* se denomina *asintóticamente R-eficiente* (a.R-e.), si

$$\mathbf{M}_\theta(\theta^* - \theta)^2 \sim \frac{1 + o(1)}{nI(\theta)} .$$

Vemos que a diferencia de las definiciones del § 8, que tenían un carácter más cualitativo, las definiciones de R-eficacia se basan en la comparación con los valores numéricos conocidos, relacionados principalmente con la información de Fisher, mejor dicho, con la cantidad $(nI(\theta))^{-1}$.

Para la R-eficacia de θ^* es necesario y suficiente el cumplimiento de (3).

De lo dicho más arriba se deduce que las estimaciones R-eficientes son eficientes, pero no al revés; las estimaciones R-eficientes simplemente existen con menos frecuencia, lo cual no es un defecto de las estimaciones, sino de la frontera inferior en la desigualdad de Rao — Cramer.

En los actuales manuales de estadística matemática, las estimaciones R-eficientes se llaman simplemente eficientes. No obstante, creemos que es más natural conservar el término «eficacia» para las mejores estimaciones en un sentido más amplio (véase la definición 8.1).

Teorema 2. *Si se han cumplido las condiciones* (R) *y existe la estimación R-eficiente, entonces esta última coincide con la estimación de verosimilitud máxima.*

Demostración. Ya hemos visto que el cumplimiento de (3) conduce a la igualdad (véase (11))

$$L'(X, \theta) = (\theta^* - \theta)c(\theta).$$

Además, como $b(\theta) = 0$, de (12) resulta

$$c(\theta) = 1/\mathbf{D}_\theta\theta^* = nI(\theta) > 0$$

para cualesquier $\theta \in \Theta$. Esto quiere decir que $L'(X, \theta) < 0$ cuando $\theta > \theta^*$, y que $L'(X, \theta) > 0$ cuando $\theta < \theta^*$. Por consiguiente, cuando $\theta = \theta^*$ se alcanza el máximo $L(X, \theta)$. ◁

El ejemplo 1 citado más arriba muestra que, a diferencia de las estimaciones R-eficientes, las estimaciones eficientes pueden no coincidir con las e.v.m. En este ejemplo, la e.v.m. es $(\overline{x})^{-1}$, mientras que la estimación eficien-

te es igual a $\frac{n-1}{n}(\bar{x})^{-1}$. Estas dos estimaciones son, evidentemente, las estimaciones a.R-e.

Examinemos la clase $\tilde{K}_0$ de las estimaciones θ^*, para las cuales, cuando $n \to \infty$,

$$|b(\theta)| \leqslant \varepsilon(\theta, n)/\sqrt{n}, \ |b'(\theta)| \leqslant \varepsilon(\theta, n),$$
$$\mathbf{M}_\theta(\theta^*)^2 < c < \infty$$

para cierta función $\varepsilon(\theta, n) = o(1)$ cuando $n \to \infty$ y cuando cada $\theta \in \Theta$.

Cada una de estas clases es notable por el hecho de que para ella la frontera inferior en la desigualdad de Rao — Cramer tiene la forma $(1 + o(1))/[nI(\theta)]$. En el § 20 veremos que en una serie de casos, al hallar las estimaciones asintóticamente óptimas, es posible limitarse al estudio de las estimaciones θ^* de tales clases.

Teorema 3. *Supongamos que se han cumplido las condiciones* (R). *Entonces, cualquier estimación a.R-e. de* $\tilde{K}_0$ *es la estimación. a.e. en* $\tilde{K}_0$.

La demostración del teorema es evidente: si θ_1^* es la estimación a.R-e., entonces

$$\mathbf{M}_\theta(\theta_1^* - \theta)^2 = \frac{1 + o(1)}{nI(\theta)}.$$

Además, como ya hemos señalado, según la desigualdad de Rao — Cramer, para todos $\theta^* \in \tilde{K}_0$,

$$\varliminf_{n \to \infty} \mathbf{M}_\theta n(\theta^* - \theta)^2 \geqslant I^{-1}(\theta) = \lim_{n \to \infty} \mathbf{M}_0 n(\theta_1^* - \theta)^2. \ \triangleleft$$

También está claro que si existe la estimación a.R-e., cualquier estimación a.e. en $\tilde{K}_0$ será la estimación a.R-e.

Más tarde (véase el § 25) veremos que con ciertas suposiciones adicionales, las estimaciones a.R-e. existen siempre y, por consiguiente, la afirmación del teorema 3 también es válida en dirección inversa: la estimación a.e. en $\tilde{K}_0$ es la estimación a.R-e. o sea, para ella $\mathbf{M}_\alpha(\theta^* - \theta)^2 \sim |nI(\theta)|^{-1}$.

Teorema 4. *Supongamos que se han cumplido las condiciones* (R). *Si* θ_1^*, θ_2^* *pertenecen a* $\tilde{K}_0$ *y son las estimaciones a.R-e., ellas son asintóticamente equivalentes en el sentido siguiente:*

$$\sqrt{n}(\theta_1^* - \theta_2^*) \underset{P}{\to} 0.$$

La demostración de esta afirmación se efectúa exactamente igual que en el teorema 8.2. Como $\theta^* = (\theta_1^* + \theta_2^*)/2 \in \tilde{K}_0$, entonces, basándonos en (8.11) y en la igualdad de Rao — Cramer, obtenemos

$$\lim_{n\to\infty} \sup \mathbf{M}_\theta n(\theta_1^* - \theta_2^*)^2 \leqslant 0. \quad \lhd$$

Ejemplo 2. La estimación $\alpha^* = \bar{x}$ del valor medio α de la población normal Φ_{α,σ^2} para σ^2 conocida es la estimación R-eficiente. Es fácil convencerse de esto, comprobando, por ejemplo, la condición (3). Otra posibilidad consiste en comparar $\mathbf{D}_\alpha\alpha^* = \sigma^2/n$ con el valor mínimo posible $(nI(\alpha))^{-1}$ de las varianzas de las estimaciones no desplazadas. En nuestro caso,

$$l(x, \alpha) = -\ln \sqrt{2\pi}\,\sigma - (x-\alpha)^2/(2\sigma^2),$$
$$l'(x, \alpha) = (x-\alpha)/\sigma^2,$$
$$I(\alpha) = \mathbf{M}_\alpha[l'(x_1, \alpha)]^2 = \mathbf{M}_\alpha(x_1-\alpha)^2/\sigma^4 = 1/\sigma^2,$$

así que $\mathbf{D}_\alpha\alpha^* = (nI(\alpha))^{-1} = \sigma^2/n$.

Ejemplo 3. Examinemos la estimación $\theta^* = S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^2$ del parámetro $\theta = \sigma^2$ de la población normal con α conocido. No es difícil calcular que $\mathbf{D}_\theta\theta^* = \mathbf{M}_\theta(\theta^* - \sigma^2)^2 = 2\sigma^4/n$. Por otro lado, aquí

$$l'(x_1, \theta) = -\frac{1}{2\theta} + \frac{(x_1-\alpha)^2}{2\theta^2},$$

$$I(\theta) = \mathbf{M}_\theta[l'(x_1, \theta)]^2 = \frac{1}{4\theta^4}\mathbf{M}_\theta[(x_1-\alpha)^2 - \theta]^2 = \frac{1}{2\theta^2} = \frac{1}{2\sigma^4}.$$

Ahora bien, aquí también $\mathbf{D}_\theta\theta^* = (nI(\theta))^{-1}$, y la estimación $\theta^* = S_1^2$ es R-eficiente.

La varianza de la estimación no desplazada $S_0^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ es igual a $\frac{2\sigma^4}{n-1}$, así que la misma no es R-eficiente o simplemente no es la estimación eficiente de σ^2. Al mismo tiempo es evidente que S_0^2 es la estimación a.R-e.

Si en calidad de parámetro desconocido estimamos no σ^2 sino $\theta = \sigma$, entonces no obtendremos la estimación R-eficiente. Sin embargo, la estimación no desplazada de σ será la estimación

$$\sigma^* = \sqrt{\frac{n}{2}}\,\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}\,S, \text{ ya que}$$

$$\mathbf{M}_\sigma S = \frac{\sigma}{\sqrt{n}}\,\mathbf{M}_\sigma\sqrt{\frac{1}{\sigma^2}\sum(x_i-\alpha)^2},$$

donde $\frac{nS^2}{\sigma^2} = \frac{1}{\sigma^2}\sum(x_i-\alpha)^2$ tiene la distribución $\mathbf{H}_n - \Gamma_{1/2,n/2}$, por eso (véase el § 2)

$$\mathbf{M}_\sigma S = \frac{\sigma\sqrt{2}}{\sqrt{n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \quad \mathbf{M}_\sigma \sigma^* = \sigma.$$

Como S es la estimación suficiente completa y mínima, σ^* es la estimación eficiente. Con ayuda de la fórmula de Stirling no es difícil convencerse de que $\sigma^* = S(1 + O(1/n))$.

Comparemos ahora la magnitud $\mathbf{D}_\sigma\sigma^*$ con la frontera inferior $(nI(\sigma))^{-1}$. Tenemos

$$\mathbf{D}_\sigma\sigma^* = \mathbf{M}_\sigma\left(\sqrt{\frac{n}{2}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} S - \sigma\right)^2 = \sigma^2\left[\frac{n}{2} \frac{\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n+1}{2}\right)} - 1\right]. \qquad (13)$$

Por otro lado, aquí

$$l'(x, \sigma) = -\frac{1}{\sigma} + \frac{(x-\alpha)^2}{\sigma^3},$$

$$I(\sigma) = \mathbf{M}_\sigma[l'(x_1, \sigma)]^2 = \frac{1}{\sigma^6}\mathbf{M}_\sigma[(x_1 - \alpha)^2 - \sigma^2]^2 = \frac{2}{\sigma^2},$$

así que $(nI(\sigma))^{-1} = \sigma^2/(2n)$. Pero este valor se distingue de (13). Su relación, por ejemplo para $n = 3$, es igual a 0,936. Ahora bien, aquí no hay estimaciones R-eficientes. Cuando $n \to \infty$ el coeficiente de σ^2 en (13) se comporta asintóticamente como $\frac{1}{2n} + O\left(\frac{1}{n^2}\right)$, así que σ^* es la estimación a.R-e.

3. Desigualdad de Rao — Cramer en el caso multidimensional. En este apartado $\theta = (\theta_1, \ldots, \theta_k)$ es el vector k-dimensional, al igual que también la estimación $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$. Como antes, pongamos

$$a(\theta) = \mathbf{M}_\theta\theta^* = \theta + b(\theta), \quad b(\theta) = (b_1(\theta), \ldots, b_k(\theta))$$

y examinemos las clases K_b de las estimaciones con un desplazamiento registrado $b(\theta)$.

La generalización de las condiciones (R) para el caso multidimensional tendrá el aspecto siguiente. Designemos

$$l(x, \theta) = \log f_\theta(x), \quad l_i(x, \theta) = \frac{\partial}{\partial\theta_i} l(x, \theta),$$

$$I_{ij}(\theta) = \mathbf{M}_\theta l_i(x_1, \theta) l_j(x_1, \theta)$$

y supongamos que se ha cumplido la condición

(R). *Las funciones $\sqrt{f_\theta(x)}$ son derivables continuamente respecto a θ_j para c.t. $[\mu]$ valores de x. La matriz*

$$I(\theta) = \|I_{ij}(\theta)\|,$$

$$I_{ij}(\theta) = \int l_i(x; \theta) l_j(x; \theta) f_\theta(x)\mu(dx)$$

es continua en θ *), *y su determinante* $|I(\theta)|$ *es distinto del cero.*

Como $I(\theta)$ es la matriz de segundos momentos $\mathbf{M}_\theta l_i l_j$ de las variables aleatorias $l_i = l_i(x_1, \theta)$, ella será una matriz definida positivamente, ya que para cualquier vector $\alpha = (\alpha_1, \ldots, \alpha_k) \neq 0$ se cumple

$$\sum \alpha_i \alpha_j \mathbf{M}_\theta l_i l_j = \mathbf{M}_\theta \left(\sum \alpha_i l_i\right)^2 \geqslant 0,$$

donde la igualdad a cero se exluye por la condición $|I(\theta)| \neq 0$.

Como antes, por desigualdad entre las matrices $\sigma_1^2 \geqslant \sigma_2^2$ entenderemos la desigualdad $\alpha\sigma_1^2\alpha^T \geqslant \alpha\sigma_2^2\alpha^T$ para cualquier vector fila $\alpha = (\alpha_1, \ldots, \alpha_k) \neq 0$. Esto equivale, evidentemente, al hecho de que la matriz $\sigma_1^2 - \sigma_2^2$ está definida de forma no negativa. La desigualdad estricta corresponderá a la definición positiva, así que, por ejemplo, $I(\theta) > 0$.

Teorema 1A. *Si* $\theta^* \in K_b$ *y si se cumple la condición* (*R*), *entonces para la matriz de segundos momentos* $\sigma^2 = |\sigma_{ij}| = \mathbf{M}_\theta(\theta^* - a(\theta))^T(\theta^* - a(\theta))$ *de cualquier estimación* θ^* *del vector fila* θ *es válida la desigualdad*

$$\sigma^2 \geqslant \frac{1}{n}(E + D(\theta))I^{-1}(\theta)(E + D(\theta))^T, \tag{14}$$

donde E es la matriz unidad, $D(\theta) = |b_{ij}(\theta)|$, $b_{ij}(\theta) = \dfrac{\partial b_i(\theta)}{\partial \theta_j}$.

Sea $|\sigma|^2 > 0$ (*o bien* $|E + D(\theta)| > 0$) *para todos* θ. *En este caso el signo de igualdad en* (14) *se alcanza si y sólo si la distribución de la muestra pertenece a una familia exponencial de tipo especial, o sea, cuando para ciertas funciones escalares* $B(\theta)$ *y* $h(X)$ *se cumple*

$$f_\theta(X) = \exp\{(\theta^*, A(\theta)) + B(\theta)\}h(X), \tag{15}$$

donde el vector $A(\theta) = (A_1(\theta), \ldots, A_k(\theta))$ *tiene una matriz de derivadas igual a*

$$|A_{ij}| = \left\| \frac{\partial A_i(\theta)}{\partial \theta_j} \right\| = n[(E + D(\theta))^{-1}]^T I(\theta).$$

Es evidente que para las estimaciones no desplazadas θ^*,

$$\sigma^2 \geqslant (nI(\theta))^{-1}$$

y la igualdad es posible únicamente cuando se cumple (15), *donde* $|A_{ij}| = nI(\theta)$.

Ahora bien, si logramos hallar la estimación no desplazada θ^* con una matriz de segundos momentos $[nI(\theta)]^{-1}$, ella será una estimación eficiente.

---

*) Para esto es suficiente exigir la convergencia uniforme de $I_{ij}(\theta)$ (véase el Suplemento VI).

En el caso multidimensional conservan su validez todas las observaciones hechas con arreglo a la desigualdad unidimensional de Rao — Cramer, así como la definición de R-eficacia, en las que deben introducirse tan sólo las modificaciones evidentes relacionadas con la dimensión de θ.

En particular, llamaremas estimaciones a.R-e. las estimaciones θ^* para las cuales

$$\mathbf{M}_\theta(\theta^* - \theta)^T(\theta^* - \theta) = \sigma^2 + b^T(\theta)b(\theta) = (nI(\theta))^{-1} + o(1/n).$$

Aquí el análogo del teorema 2 tendrá el aspecto siguiente.

Teorema 2A. *Supongamos que se cumplen las condiciones* (R). *Si* θ^* *es la estimación R-eficiente, entonces ésta es la estimación de verosimilitud máxima.*

Demostración. Para demostrar que la estimación R-eficiente constituye el único punto del máximo, es suficiente convencerse que $L'(X, \theta^*) = 0$ y que cuando $\theta = \theta^* + u$, $u \neq 0$,

$$(\text{grad}\, L(X, \theta), u) = (L'(X, \theta), \theta - \theta^*) < 0.$$

Pero en el caso de existencia de la estimación R-eficiente, se cumple (véase (20))

$$L'(X, \theta) = (\theta^* - \theta)nI(\theta),$$

de donde se desprenden inmediatamente las relaciones requeridas. La segunda se deduce del hecho de que

$$(L', u) = -unI(\theta)u^T,$$

donde $uI(\theta)u^T$ es la forma cuadrática definida positivamente. ◁

Ejemplo 4. Examinemos una familia biparamétrica de distribuciones normales Φ_{α,σ^2}. La misma pertenece a una familia exponencial, ya que (aquí $\theta = (\theta_1, \theta_2)$, $\theta_1 = \alpha$, $\theta_2 = \sigma^2$)

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\alpha)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2} + \frac{x\alpha}{\sigma^2} - \frac{\alpha^2}{2\sigma^2} - \ln\sigma\right\}.$$

La estimación $\theta^* = (\theta_1^*, \theta_2^*)$, donde $\theta_1^* = \bar{x}$, $\theta_2^* = S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum x_i^2 - \bar{x}^2\right)$ es eficiente, puesto que pertenece a K_0, y la estadística $\left(\sum x_i, \sum x_i^2\right)$, como hemos visto en el § 15, es la estadística suficiente completa (véase el teorema 14.4).

Señalemos que

$$\mathbf{M}_\theta(\theta^* - \theta)^T(\theta^* - \theta) = \sigma^2 + b^T(\theta)b(\theta).$$

Demostración del teorema 1A. Designemos

$$L_j' = L_j'(X, \theta) = \sum_{i=1}^{n} l_j'(x_i, \theta), \; L' = L'(X, \theta) = (L_1', \ldots, L_k').$$

Entonces, de un modo completamente análogo al caso unidimensional, establecemos que son válidas las igualdades

$$\mathbf{M}_\theta l_j'(x_1, \theta) = 0, \; \mathbf{M}_\theta \theta_i^* L_j'(X, \theta) = 1 + b_{ij}(\theta)$$

en las cuales $b_{ij}(\theta)$ son continuas o bien, que es lo mismo, las igualdades

$$\mathbf{M}_\theta L' = 0, \tag{16}$$

$$\mathbf{M}_\theta(\theta^*)^T L' = E + D(\theta) \tag{17}$$

en las que la matriz $D(\theta)$ es continua. De aquí obtenemos

$$\mathbf{M}_\theta(\theta^* - a(\theta))^T L' = E + D(\theta). \tag{18}$$

Demostremos ahora la desigualdad siguiente (variante matricial de la desigualdad de Cauchy — Buniakovski).

Lema 2. *Supongamos que ξ y η son matrices de igual dimensión (no obligatoriamente cuadradas) con elementos aleatorios, y que la matriz $\mathbf{M}\eta\eta^T$ tiene inversa. Entonces*

$$\mathbf{M}\xi\xi^T \geqslant \mathbf{M}\xi\eta^T(\mathbf{M}\eta\eta^T)^{-1}\mathbf{M}\eta\xi^T. \tag{19}$$

En este caso la igualdad es posible únicamente cuando $\xi = z\eta$, $z = \mathbf{M}\xi\eta^T(\mathbf{M}\eta\eta^T)^{-1}$.

Demostración. En vista de que para cualquier matriz A es válida la desigualdad $AA^T \geqslant 0$ (AA^T está definida no negativamente), entonces

$$0 \leqslant \mathbf{M}(\xi - z\eta)(\xi - z\eta)^T = \mathbf{M}\xi\xi^T - z\mathbf{M}\eta\xi^T - \mathbf{M}\xi\eta^T z^T + z\mathbf{M}\eta\eta^T z^T.$$

Poniendo $z = \mathbf{M}\xi\eta^T(\mathbf{M}\eta\eta^T)^{-1}$, obtenemos la desigualdad requerida.

La afirmación con respecto a las condiciones de la igualdad en (19) es evidente. ◁

Volvamos a la demostración del teorema 1A. Pongamos, en (19), $\xi = (\theta^* - a(\theta))^T$, $\eta = (L')^T$. Entonces

$$\mathbf{M}_\theta\xi\xi^T = \mathbf{M}_\theta(\theta^* - a(\theta))^T(\theta^* - a(\theta) = \sigma^2.$$

De (16) y de la desigualdad de x_i obtenemos

$$\mathbf{M}_\theta\eta\eta^T = \mathbf{M}_\theta(L')^T L' = nI(\theta).$$

Por último, de (18) hallamos

$$\mathbf{M}_\theta \xi\eta^T = \mathbf{M}_\theta(\theta^* - a(\theta))^T L' = E + D(\theta).$$

La desigualdad (14) queda demostrada.

La desigualdad en (14) es posible en virtud del lema 2, si sólo para los puntos (x, θ), tales que $f_\theta(x) > 0$, es válida

$$(\theta^* - a(\theta))^T = (E + D(\theta))(nI(\theta)^{-1}(L')^T$$

o, que es lo mismo,

$$L' = (\theta^* - a(\theta))n[(E + D(\theta))^{-1}]^T I(\theta). \qquad (20)$$

Nótese ahora que de la desigualdad en (14) resulta

$$|E + D(\theta)|^2 = n|\sigma^2|\cdot|I(\theta)|,$$

y la separación del determinante $|\sigma^2|$ de 0 quiere decir lo mismo para $|E + D(\theta)|$ y significa la existencia de la matriz inversa $(E + D(\theta))^{-1}$ uniformemente limitada. Por eso la derivada L' en (20) será limitada, y $f_\theta(x) > 0$ en todas partes de Θ y la misma igualdad (20) será válida en todas partes de Θ. Si ahora s es cualquier camino que une los puntos θ_1 y θ en la región Θ, entonces

$$L(X, \theta) = \int_s (L', ds) + L(X, \theta_0),$$

donde ds significa el elemento vectorial del camino s; $((L', ds) = (L', s'(l))dl$ es el incremento $L(X, \theta)$ en dicho camino; y l, la «longitud» del camino recorrido. Por consiguiente, en virtud de (20),

$$L(X, \theta) = \theta^* A(\theta) + B(\theta) + H(X), \qquad (21)$$

donde $B(\theta)$ y $H(X)$ son funciones escalares; $A(\theta) = (A_1(\theta), \ldots, A_k(\theta))$ es un vector que depende exclusivamente de sus argumentos. Esto significa la validez de (15).

Si se cumple (21), entonces

$$L' = \theta^* |A_{ij}| + B'(\theta),$$

donde, en virtud de la igualdad $\mathbf{M}_\theta L' = 0$, es válida

$$B'(\theta) = -a(\theta)|A_{ij}|.$$

Multiplicando ambos miembros de la igualdad $L' = (\theta^* - a(\theta))|A_{ij}|$, a la izquierda en $(\theta^* - a(\theta))^T$, obtenemos, en virtud de (18), que para el cumplimiento de la condición (20), que significa la igualdad en (14), debe cumplirse

$$|A_{ij}| = n[(E + D(\theta))^{-1}]^T I(\theta). \quad \triangleleft$$

En el caso multidimensional conservan su validez todas las observaciones hechas con arreglo a la desigualdad unidimensional de Rao — Cramer, así como la definición de R-eficacia, en las que deben introducirse tan sólo las modificaciones evidentes relacionadas con la dimensión de θ.

En particular, llamaremas estimaciones a.R-e. las estimaciones θ^* para las cuales

$$\mathbf{M}_\theta(\theta^* - \theta)^T(\theta^* - \theta) = \sigma^2 + b^T(\theta)b(\theta) = (nI(\theta))^{-1} + o(1/n).$$

Aquí el análogo del teorema 2 tendrá el aspecto siguiente.

Teorema 2A. *Supongamos que se cumplen las condiciones* (R). *Si* θ^* *es la estimación R-eficiente, entonces ésta es la estimación de verosimilitud máxima.*

Demostración. Para demostrar que la estimación R-eficiente constituye el único punto del máximo, es suficiente convencerse que $L'(X, \theta^*) = 0$ y que cuando $\theta = \theta^* + u$, $u \neq 0$,

$$(\operatorname{grad} L(X, \theta), u) = (L'(X, \theta), \theta - \theta^*) < 0.$$

Pero en el caso de existencia de la estimación R-eficiente, se cumple (véase (20))

$$L'(X, \theta) = (\theta^* - \theta)nI(\theta),$$

de donde se desprenden inmediatamente las relaciones requeridas. La segunda se deduce del hecho de que

$$(L', u) = -unI(\theta)u^T,$$

donde $uI(\theta)u^T$ es la forma cuadrática definida positivamente. ◁

Ejemplo 4. Examinemos una familia biparamétrica de distribuciones normales Φ_{α,σ^2}. La misma pertenece a una familia exponencial, ya que (aquí $\theta = (\theta_1, \theta_2)$, $\theta_1 = \alpha$, $\theta_2 = \sigma^2$)

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\alpha)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2} + \frac{x\alpha}{\sigma^2} - \frac{\alpha^2}{2\sigma^2} - \ln\sigma\right\}.$$

La estimación $\theta^* = (\theta_1^*, \theta_2^*)$, donde $\theta_1^* = \bar{x}$, $\theta_2^* = S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum x_i^2 - \bar{x}^2\right)$ es eficiente, puesto que pertenece a K_0, y la estadística $\left(\sum x_i, \sum x_i^2\right)$, como hemos visto en el § 15, es la estadística suficiente completa (véase el teorema 14.4).

La estimación de verosimilitud máxima $\left(\bar{x}, \frac{1}{n}\sum(x_i - \bar{x})^2\right)$ se distingue de θ^* sólo por el factor $\frac{n-1}{1}$ de la segunda coordenada, debido a lo cual la misma permanece desplazada. Para la estimación elegida θ^*, la representación exponencial especial (15) de la función $f_\theta(X)$ no se realizará, ya que

$$f_\theta(X) = (2\pi)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum x_i^2 + \frac{\alpha}{\sigma^2}\sum x_i - \frac{n\alpha^2}{2\sigma^2} - n\ln\sigma\right\} =$$

$$=(2\pi)^{-n/2}\exp\left\{\frac{\alpha n}{\sigma^2}\theta_1^* - \frac{n-1}{2\sigma^2}\theta_2^* - \frac{n}{2\sigma^2}(\theta_1^*)^2 - \frac{n\alpha^2}{2\sigma^2} - n\ln\sigma\right\}.$$

Esto significa que en la desigualdad multidimensional de Rao — Cramer no será alcanzada la frontera inferior.

El elipsoide de dispersión mínimo, definido (según el teorema 1*A*) por la matriz $I(\theta)$ (o $I^{-1}(\theta)$), se alcanzará sólo asintóticamente cuando $n \to \infty$, así que la estimación θ^*, sin ser *R*-eficiente, será la estimación a.*R*-e. Cerciorémonos de ello directamente.

Calculemos al principio la matriz $I(\theta)$. Tenemos

$$l_1'(x, \theta) = \frac{(x-\alpha)}{\sigma^2}, \quad l_2'(x, \theta) = \frac{(x-\alpha)^2}{2\sigma^4} - \frac{1}{2\sigma^2}$$

(recordemos que l_2' no es derivada respecto a σ sino respecto a σ^2, comparen esto con el ejemplo 3). Por eso

$$I_{11}(\theta) = \mathbf{M}_\theta \frac{(x_1-\alpha)^2}{\sigma^4} = \frac{1}{\sigma^2},$$

$$I_{12}(\theta) = I_{21}(\theta) = \mathbf{M}_\theta\left[\frac{(x_1-\alpha)^3}{2\sigma^6} - \frac{x_1-\alpha}{2\sigma^4}\right] = 0,$$

$$I_{22}(\theta) = \frac{1}{4\sigma^8}\mathbf{M}_\theta[(x_1-\alpha)^2 - \sigma^2]^2 = \frac{1}{2\sigma^4}.$$

De aquí hallamos

$$(nI(\theta))^{-1} = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}. \qquad (22)$$

Calculemos ahora, para comparar, la matriz de segundos momentos centrales de la estimación θ^*.

Tenemos

$$\mathbf{M}_\theta(\theta_1^* - \theta_1)^2 = \mathbf{M}_\theta(\bar{x} - \alpha)^2 = \frac{\sigma^2}{n},$$

$$\mathbf{M}_\theta(\theta_2^* - \theta_2)^2 = \mathbf{M}_\theta(S_0^2 - \sigma^2)^2 = \frac{2\sigma^4}{n-1},$$

$$\mathbf{M}_\theta(\theta_1^* - \theta_1)(\theta_2^* - \theta_2) = 0.$$

Las dos últimas ecuaciones se calculan directamente. Examinemos, por ejemplo, la segunda de ellas. Es suficiente convencernos de que

$$\mathbf{M}_\theta(\bar{x} - \alpha)S_0^2 = 0. \tag{23}$$

Pero

$$S_0^2 = \frac{1}{n-1}\left[\sum(x_i - \alpha)^2 - (\bar{x} - \alpha)^2\right],$$

$$(\bar{x} - \alpha)S_0^2 = \frac{1}{n(n-1)}\left[\sum(x_i - \alpha)\right]\left[\sum(x_i - \alpha)^2\right] - \frac{1}{n(n-1)}(\bar{x} - \alpha)^3.$$

En vista de que

$$\mathbf{M}_\theta(\bar{x} - \alpha)^3 = \mathbf{M}_\theta(x_i - \alpha)^3 = \mathbf{M}_\theta(x_i - \alpha)(x_j - \alpha)^2 = 0,$$

(23) queda demostrada.

Ahora bien, la matriz de segundos momentos $\theta^* - \theta$ es igual a

$$\begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/(n-1) \end{bmatrix}.$$

Por supuesto que la diferencia entre esta matriz y la matriz $(nI(\theta))^{-1}$ puede ser considerable sólo para pequeños valores de n.

4. Algunas deducciones. Concluyendo este párrafo, hagamos cierto resumen de las investigaciones realizadas en los seis últimos párrafos. Su finalidad principal consistía en buscar los métodos de construir las estimaciones óptimas (en uno u otro sentido) y fijar las fronteras inferiores para sus desviaciones estándar. Como resultado se pueden indicar las siguientes cuatro tendencias principales de búsqueda de las mejores estimaciones.

1. Construcción de las estimaciones bayesianas (si hay una información a priori sobre θ) y minimax.

2. Determinación de las estadísticas suficientes completas (o mínimas) S. Entonces la estimación $\theta_S^* = \mathbf{M}_\theta(\theta^*/S)$ será eficiente en la clase $\mathbf{K}_b$, a la cual pertenece θ^*.

3. Utilización de las e.v.m. en los casos en que se cumple el criterio (3) del teorema 1 (o el criterio (15) del teorema 1A). En este caso también obtendremos las estimaciones eficientes (e incluso R-eficientes) en las clases con un desplazamiento registrado.

4. Enfoque cuantitativo basado en la comparación de la desviación estándar $\mathbf{M}_\theta(\theta^* - \theta)^2$ de la estimación θ^*, que queremos utilizarla, con la frontera inferior R definida por la desigualdad de Rao — Cramer. Si la relación $\mathbf{M}_\theta(\theta^* - \theta)^2/R$ es próxima a cero, la estimación θ^* puede ser recomendada para el uso. Siguiendo esta tendencia, obtendremos ulteriormente resultados muy generales relacionados con la construcción de las estimacio-

nes asintóticamente eficientes, asintóticamente bayesianas y asintóticamente minimax.

Hagamos también la siguiente observación. En todas las tendencias señaladas más arriba, desempeña un papel muy importante la forma en que la distribución de la muestra $\mathbf{P}_\theta$ depende del parámetro θ que se estima. Sin embargo, en la práctica a menudo surgen problemas de no estimación del propio θ sino de cierta función $\varphi(\theta)$ de éste. Además es facil notar (véase el ejemplo con el esquema de Bernoulli en (8.4) y (8.5)) que la estimación $\varphi^* = \varphi(\theta^*)$ no siempre, ni mucho menos, poseerá las propiedades que poseía la estimación θ^* (no estar desplazada, ser eficaz, etc., sólo se conservarán las propiedades de eficacia asintótica si φ es una función suave). Desde este punto de vista es natural que al principio se examine el problema de estimación de las *funciones* $\varphi(\theta)$ del parámetro inicial θ. Pero hemos renunciado a tal enfoque, ya que, manteniendo esta tendencia, muchos resultados básicos, obtenidos por nosotros, se complicarían considerablemente. Por otro lado, si φ realiza una aplicación biunívoca, el problema de estimación de $\varphi(\theta)$ se reducirá al problema examinado por nosotros mediante la «reparametrización», o sea, la introducción de un nuevo parámetro $\gamma = \varphi(\theta)$, al que le corresponderá la familia de distribuciones $\mathbf{G}_\gamma = \mathbf{P}_{\varphi^{-1}(\gamma)}$.

§ 17*. Propiedades de la información de Fisher

Ya hemos visto, y nos convenceremos en adelante, que la información de Fisher desempeña un papel muy importante en la estadística matemática. Por eso aclaremos algunas propiedades útiles de la misma.

1. Caso unidimensional. La información de Fisher,

$$I(\theta) = \int \frac{(f'_\theta(x))^2}{f_\theta(x)}\,\mu(dx) = \mathbf{M}_\theta(l'(\mathrm{x}_1,\ \theta))^2,$$

apareció en las investigaciones del párrafo precedente. La magnitud

$$I^X(\theta) = \mathbf{M}_\theta[L'(X,\ \theta)]^2$$

suele considerarse como la *medida de la cantidad de información contenida en la muestra X respecto al parámetro θ*. En el teorema 16.1 hemos demostrado la *aditividad* de la información: $I^n(\theta) = nI(\theta)$, o sea, que $I^n(\theta)$ es igual a la suma de informaciones $I^{\mathrm{x}_i}(\theta) = \mathbf{M}_\theta[l'(\mathrm{x}_i,\ \theta)]^2 = I(\theta)$ contenidas en las observaciones independientes $\mathrm{x}_1, \ldots, \mathrm{x}_n$.

Demostremos una propiedad más de la información de Fisher. Sea $S = S(X)$ cierta estadística con valores en R^l, y sea $g_\theta(s)$ la densidad de su distribución inducida por la distribución $\mathbf{P}_\theta$ en $(\mathscr{X}^n,\ \mathfrak{B}_{\mathscr{X}}^n)$ respecto a cierta medida λ en $(R^l,\ \mathfrak{B}^l)$. De acuerdo con las designaciones anteriores,

llamaremos la magnitud

$$I^S(\theta) = \mathbf{M}_\theta[(\log g_\theta(S))']^2$$

información contenida en la estadística S respecto al parámetro θ.

Notemos que el valor de $I^s(\theta)$ no depende de la elección de la medida λ. En efecto, si $\tilde{\lambda}$ es cualquier otra medida y $\nu = \lambda + \tilde{\lambda}$. Entonces λ y $\tilde{\lambda}$ serán absolutamente continuas respecto a ν, y la densidad $g_\theta^\nu(s)$ de la distribución de S respecto a la medida ν será igual a

$$g_\theta^\nu(s) = g_\theta(s)\, \frac{d\lambda}{d\nu} = \tilde{g}_\theta(s) \frac{d\tilde{\lambda}}{d\nu},$$

donde $\tilde{g}_\theta$ es la densidad respecto a $\tilde{\lambda}$. Como $\frac{d\lambda}{d\nu}$ y $\frac{d\tilde{\lambda}}{d\nu}$ no dependen de θ, las derivadas de los logaritmos de todas las tres expresiones coincidirán.

Teorema 1. *Supongamos que las densidades $f_\theta(x)$ y $g_\theta(s)$ satisfacen las condiciones (R). Entonces*

$$I^s(\theta) \leqslant I^x(\theta). \tag{1}$$

Aquí la igualdad se alcanza si y sólo si S es una estadística suficiente.

Demostración. Para cualquier $B \in \mathfrak{B}^l$ designemos por $S^{-1}(B) \in \mathfrak{B}^n_{\mathscr{X}}$ el conjunto $x \in \mathscr{X}^n$ para el cual $S(x) \in B$. Entonces según la definición de la e.m.c.,

$$\int_{S^{-1}(B)} L'(x, \theta)\mathbf{P}_\theta(dx) = \mathbf{M}_\theta[L'(X, \theta);\ X \in S^{-1}(B)] =$$
$$= \mathbf{M}_\theta[\mathbf{M}_\theta(L'(X, \theta)/S);\ S \in B]. \tag{2}$$

Por otro lado,

$$\int_{S^{-1}(B)} L'(x, \theta)\mathbf{P}_\theta(dx) = \frac{\partial}{\partial\theta} \int_{S^{-1}(B)} f_\theta(x)\mu^n(dx) = \frac{\partial}{\partial\theta} \int_B g_\theta(s) \times$$
$$\times \lambda(ds) = \int_B \frac{\partial}{\partial\theta}\, g_\theta(s)\lambda(ds) = \mathbf{M}_\theta[(\log g_\theta(S))';\ S \in B]. \tag{3}$$

Comparando (2) y (3), vemos que c.d. $[\mathbf{P}_\theta]$

$$\mathbf{M}_\theta(L'(X, \theta)/S) = (\log g_\theta(S))'. \tag{4}$$

Luego tenemos

$$0 \leqslant \mathbf{M}_\theta[L'(X, \theta) - (\log g_\theta(S))']^2 =$$
$$= I^X(\theta) + I^S(\theta) - 2\mathbf{M}_\theta L'(X, \theta)(\log g_\theta(S))',$$

donde, en virtud de (4),

$$\mathbf{M}_\theta L'(X,\ \theta)(\log g_\theta(S))' =$$
$$= \mathbf{M}_\theta[(\log g_\theta(S))'\mathbf{M}_\theta(L'(X,\ \theta)/S] = \mathbf{M}_\theta[(\log g_\theta(S))']^2 = I^S(\theta).$$

Esto demuestra la desigualdad (1).

Sea ahora S una estadística suficiente para θ. Entonces

$$f_\theta(X) = \psi(S,\ \theta)h(X). \tag{5}$$

Tomemos en calidad de λ la medida

$$\lambda(B) = \int\limits_{S^{-1}(B)} h(x)\mu^n(dx).$$

Entonces, como se muestra en el lema 15.1, la distribución de S será absolutamente continua respecto a λ y tendrá una densidad $g_\theta(s)$ igual a $g_\theta(s) = \psi(s,\ \theta)$. De aquí, en virtud de (5), obtenemos

$$I^X(\theta) = \mathbf{M}[L'(X,\ \theta)]^2 = \mathbf{M}_\theta[(\log \psi(S,\ \theta))']^2 = Y^S(\theta).$$

Mostremos ahora que de todas las igualdades $Y^X(\theta) = Y^S(\theta)$ para todos θ se deduce que S es estadística suficiente. Efectivamente, $Y^X(\theta)$ es la dispersión de $L'(X,\ \theta)$, así que

$$I^X(\theta) = \mathbf{M}_\theta[L'(X,\ \theta) - \mathbf{M}_\theta(L'(X,\ \theta)/S)]^2 + \mathbf{M}_\theta[\mathbf{M}_\theta(L'(X,\ \theta)/S]^2. \tag{6}$$

Pero, en virtud de (4), el último sumando es igual a

$$\mathbf{M}_\theta[(\log g_\theta(S))']^2 = I^S(\theta).$$

Como $I^X(\theta) = I^S(\theta)$, entonces en (6) c.d. $[\mathbf{P}_\theta]$ para todos θ,

$$L'(X,\ \theta) - \mathbf{M}_\theta(L'(X,\ \theta)/S) = 0.$$

Por lo tanto, $L'(X,\ \theta)$ es medible respecto a $\sigma(S)$ y, por consiguiente, existe una función medible $\varphi(S,\ \theta)$ tal que

$$L'(X,\ \theta) = \varphi(S,\ \theta),\ L(X,\ \theta) = \Phi(S,\ \theta) + h_1(X),$$
$$f_\theta(X) = \exp[\Phi(S,\ \theta) + h_1(X)].\ \lhd$$

Ya hemos señalado que las estadísticas suficientes son el tipo único de estadísticas que reducen los datos muestrales sin perder la información acerca del parámetro θ. El teorema 1 confiere a esta afirmación el sentido exacto con arreglo a la información de Fisher.

Ejemplo 1. Sea $X \in \mathbf{B}_p$. Aquí

$$f_p(x) = p^x(1-p)^{1-x},$$

donde x es igual a 0 ó a 1, y $f_p(x)$ es la densidad respecto a la medida

de cálculo. Por eso

$$l(x, p) = x \ln p + (1 - x)\ln(1 - p),$$
$$l'(x, p) = \frac{x}{p} - \frac{1 - x}{1 - p},$$
$$I(p) = \mathbf{M}_p[l'(x_1, p)]^2 = p\left(\frac{1}{p}\right)^2 + (1 - p)\left(\frac{1}{1 - p}\right)^2 = \frac{1}{p(1 - p)}.$$

Ahora bien, la información de una observación en el esquema de Bernoulli es igual a $(p(1 - p))^{-1}$ y alcanza su valor mínimo cuando $p = 1/2$.

La información de toda la muestra constituye $n/(p(1 - p))$. Designemos ahora por ν el número de «casos favorables» en la muestra X (número de casos unitarios) y hallemos la información de esta observación. Las densidades (otra vez respecto a la medida de cálculo) para ν serán iguales a

$$g_p(x) = C_n^x p^x (1 - p)^{n - x}, \; x = 0, \ldots, n,$$

así que $\log g_p(x) = x \log p + (n - x)\log(1 - p) + \log C_n^x$,

$$I^{\nu}(p) = \mathbf{M}_p[(\log g_p(\nu))']^2 =$$
$$= \sum_{x=0}^{n} C_n^x p^x (1 - p)^{n - x}\left(\frac{x}{p} - \frac{n - x}{1 - p}\right)^2 = \sum_{x=0}^{n} C_n^x p^x (1 - p)^{n - x} \times$$
$$\times \frac{(x - np)^2}{(p(1 - p))^2} = \frac{1}{(p(1 - p))^2}\,\mathbf{D}\nu = \frac{n}{p(1 - p)}.$$

Esta igualdad concuerda por completo con el teorema 1.

Le proponemos al lector que halle, en forma de ejercicios, las informaciones de observaciones para las muestras de las distribuciones que dependen del parámetro unidimensional y que han sido dadas en el § 2.

2. Caso multidimensional. Sea ahora $\theta \in R^k$, $k > 1$. En este caso se trata de la matriz de información de Fisher de la observación x_1:

$$I(\theta) = \|I_{ij}(\theta)\|, \; I_{ij}(\theta) = \mathbf{M}_\theta \frac{\partial}{\partial \theta_i} l(x_1, \theta) \frac{\partial}{\partial \theta_j} l(x_1, \theta),$$

donde se supone, claro está, que la función $f_\theta(x)$ es derivable.

Si ponemos

$$\varphi(x, \theta) = (\varphi_1(x, \theta), \ldots, \varphi_k(x, \theta)) =$$
$$= 2(\sqrt{f_\theta(x)})' = \frac{1}{\sqrt{f_\theta(x)}}\left(\frac{\partial f_\theta(x)}{\partial \theta_1}, \ldots, \frac{\partial f_\theta(x)}{\partial \theta_k}\right),$$

entonces la matriz $I(\theta)$ también puede ser escrita en la forma

$$I(\theta) = \int_{\mathscr{X}} \varphi^T(x, \theta)\varphi(x, \theta)\mu(dx).$$

Ya hemos establecido, en el § 16, que al igual que en el caso unidimensional, la información de Fisher es aditiva, o sea, la matriz de información de Fisher de la muestra X es igual a la suma de las matrices de información de distintas observaciones. Si designamos

$$I^X(\theta) = \|I^X_{ij}(\theta)\|, \quad I^X_{ij}(\theta) = \mathbf{M}_\theta \frac{\partial}{\partial \theta_i} L(X, \theta) \frac{\partial}{\partial \theta_j} L(X, \theta),$$

entonces $I^X(\theta) = nI(\theta)$.

El teorema 1 también es completamente válido. Sea $g_\theta(s)$ la densidad de cierta estadística $S = S(X)$ con valores en R^l respecto a cierta medida λ. Designemos

$$I^S(\theta) = \|I^S_{ij}(\theta)\|, \quad I^S_{ij}(\theta) = \mathbf{M}_\theta \frac{\partial}{\partial \theta_i} \log g_\theta(S) \frac{\partial}{\partial \theta_j} \log g_\theta(S).$$

Hemos obtenido la matriz de información de la observación S.

Teorema 1A. *Si las densidades $f_\theta(x)$ y $g_\theta(s)$ satisfacen las condiciones (R) del § 16, entonces*

$$I^S(\theta) \leqslant I^X(\theta), \tag{7}$$

o sea, la matriz $I^X(\theta) - I^S(\theta)$ es definida no negativamente. La igualdad en (7) tiene lugar si y sólo si S es una estadística suficiente.

La demostración de este teorema es completamente análoga a la del teorema 1 y, para abreviar, la omitimos. La misma se puede hallar, por ejemplo, en [95] y [48].

Ejemplo 2. En el § 16 ya hemos calculado la matriz de información para una distribución normal. Calculémosla ahora para una familia biparamétrica de distribuciones

$$f_\theta(x) = \frac{1}{\sigma} f\left(\frac{x - \alpha}{\sigma}\right),$$

donde $\theta = (\alpha, \sigma)$, f es una función derivable dada, para la cual existen las integrales

$$I_i = \int x^i \frac{(f'(x))^2}{f(x)} dx = \mathbf{M}_{(0,1)} x_1^i (l'(x_1))^2, \quad i = 0, 1, 2.$$

Aquí $l(x) = \log f(x)$; la tilde $'$ significa la derivación ordinaria, y α y σ son los parámetros de desplazamiento y escala de una distribución de densidad $f(x)$. Ahora bien, conocemos el tipo de la distribución, pero sólo con una exactitud de hasta la transformación lineal del argumento. Los parámetros α y σ de la distribución normal Φ_{α,σ^2} son, evidentemente, los parámetros de desplazamiento y escala. Al ser registrado λ, el parámetro λ de la

distribución Γ es un parámetro de escala, al igual que el parámetro θ en la distribución $U_{0,\theta}$.

Tenemos

$$l(x, \theta) = \log f_\theta(x) = -\log\sigma + l\left(\frac{x-\alpha}{\sigma}\right),$$

$$\frac{\partial l(x, \theta)}{\partial\alpha} = -\frac{1}{\sigma} l'\left(\frac{x-\alpha}{\sigma}\right),$$

$$\frac{\partial l(x, \theta)}{\partial\sigma} = -\frac{1}{\sigma} - \frac{(x-\alpha)}{\sigma^2} l'\left(\frac{x-\alpha}{\sigma}\right).$$

De aquí hallamos

$$I_{11}(\theta) = \frac{1}{\sigma^2} M_\theta\left[l'\left(\frac{x_1-\alpha}{\sigma}\right)\right]^2 = \frac{1}{\sigma^2}\int \frac{\left[f'\left(\frac{x-\alpha}{\sigma}\right)\right]^2}{\sigma f\left(\frac{x-\alpha}{\sigma}\right)}dx = \frac{1}{\sigma^2} I_0,$$

$$I_{12}(\theta) = \frac{1}{\sigma^2} M_\theta l'\left(\frac{x_1-\alpha}{\sigma}\right)\left[1 + \frac{x_1-\alpha}{\sigma} l'\left(\frac{x_1-\alpha}{\sigma}\right)\right] = \frac{1}{\sigma^2} I_1,$$

$$I_{22}(\theta) = \frac{1}{\sigma^2} M_\theta\left[1 + \frac{x_1-\alpha}{\sigma} l'\left(\frac{x_1-\alpha}{\sigma}\right)\right]^2 = \frac{1}{\sigma^2}[I_2 - 1],$$

puesto que $2\int\frac{x-\alpha}{\sigma} f'\left(\frac{x-\alpha}{\sigma}\right)\frac{dx}{\sigma} = -2\int f(x)dx = -2$. Por lo tanto,

$$I(\theta) = \frac{1}{\sigma^2}\left\|\begin{matrix} I_0 & I_1 \\ I_1 & I_2 - 1\end{matrix}\right\|.$$

Si f es una función simétrica, es evidente que $I_1 = 0$.

La degeneración de la matriz $I(\theta)$ significa que su determinante se reduce a cero o, que es lo mismo,

$$[M_{(0,1)} l'(x_1)(1 + x_1 l'(x_1))]^2 = M_{(0,1)}(l'(x_1))^2 M_{(0,1)}(1 + x_1 l'(x_1))^2.$$

Esto es posible únicamente en el caso cuando $1 + xl'(x) = cl'(x)$ para cualquier c, o cuando $l'(x) = 0$. De la primera igualdad se deduce que

$$l(x) = -\ln(x-c) + c_1,\ f(x) = \frac{e^{c_1}}{x-c}.$$

Está claro que tal función $f(x)$ no puede ser la densidad de la distribución. Análogamente se examina la posibilidad de que $l'(x) = 0$. Por lo tanto, $I(\theta)$ está definida positivamente.

En particular, para la familia normal $\{\Phi_{\alpha,\sigma^2}\}$, cuando $\theta = (\alpha, \sigma)$,

$$I(\theta) = \frac{1}{\sigma^2}\begin{Vmatrix} 1 & 0 \\ 0 & 2 \end{Vmatrix},$$

puesto que en este caso $l(x) = -x^2/2 - \ln\sqrt{2\pi}$, $l'(x) = -x$, $I_0 = M_{(0,1)}x_1^2 = 1$, $I_1 = M_{(0,1)}x_1^3 = 0$, $I_2 = M_{(0,1)}x_1^4 = 3$. Podríamos haber obtenido este mismo resultado con ayuda del ejemplo 16.4, si hubiéramos utilizado los datos del apartado 3 donde hemos mostrado el comportamiento de la matriz de información al sustituir el parámetro (en el ejemplo 16.4 $\theta = (\alpha, \sigma^2)$, pero no (α, σ)). Le proponemos al lector que se cerciore de que, en concordancia con el teorema 1A, la estadística $\left(\bar{x}, \sum x_i^2\right)$ tiene la matriz de información

$$I^S(\theta) = \frac{1}{\sigma^2}\begin{Vmatrix} 1 & 0 \\ 0 & 2 \end{Vmatrix} = nI(\theta).$$

3. Matriz de Fisher y sustitución del parámetro. Examinemos la cuestión de cómo se comporta la matriz de información al sustituir el parámetro. Pongamos $\theta = v(\beta)$, $\beta \in R^k$, donde v es una función vectorial derivable, y examinemos la familia paramétrica $\mathbf{P}_\beta^{(1)} = \mathbf{P}_{v(\beta)}$. Con el fin de hallar la matriz de información $J(\beta)$ para esta familia, debemos hallar las derivadas

$$\frac{\partial}{\partial \beta_j} l(x_1, v(\beta)) = \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} l(x_1, v(\beta)) \frac{\partial v_i(\beta)}{\partial \beta_j}. \tag{8}$$

Si designamos $V = \begin{Vmatrix} \dfrac{\partial v_i(\beta)}{\partial \beta_j} \end{Vmatrix}$, $i, j = 1, \ldots, k$, obtenemos que el vector de las derivables en (8) $l'_\beta(x_1, v(\beta))$ es representable en la forma $l'_\theta(x_1, v(\beta))V$, así que

$$J(\beta) = M_\beta(l'_\theta(x_1, v(\beta))V)^T(l'_\theta(x_1, v(\beta))V) = V^T I(v(\beta))V.$$

En particular, si $\theta = \beta C$, $C = \|c_{ij}\|$, $i, j, = 1, \ldots, k$, entonces $V = C^T$ y

$$J(\beta) = CI(\theta)C^T. \tag{9}$$

Obsérvese que si examinamos, en el espacio paramétrico, el elipsoide

$$(\theta - \theta_1)I(\theta)(\theta - \theta_1)^T < c, \tag{10}$$

la escritura (10) de este conjunto es invariante con respecto a la transformación invertible lineal C sobre el parámetro θ. Así pues, si ponemos $\theta = \beta C$, el conjunto (10) en nuevas variables tendrá la forma

$$(\beta - \beta_1)J(\beta)(\beta - \beta_1)^T < c,$$

donde $\beta_1 = \theta_1 C^{-1}$. Esto se obtiene inmediatamente si se sustituye $\theta = \beta C$ en (10) y si utilizamos (9).

§ 18*. Estimaciones del parámetro de desplazamiento y escala. Estimaciones equivariantes eficientes

En los §§ 12—16 hemos visto y nos convenceremos posteriormente hasta qué punto es útil el concepto de estadística suficiente en general y al construir las estimaciones eficientes en particular. El círculo de ideas relacionadas con la utilización de las estadísticas suficientes podría llamarse *principio de suficiencia*. Al construir las estimaciones eficientes hemos combinado el principio de suficiencia con otro principio llamado *principio de no desplazamiento*. Este último consiste en separar las clases de estimaciones con desplazamento registrado y, en particular, con desplazamiento nulo. Sin registrar el desplazamiento sería imposible separar las estimaciones eficientes.

En este párrafo, así como en el párrafo siguiente y en el capítulo 3, examinaremos el tercer principio importante de la estadística matemática, o sea, el *principio de invariación*.

La introducción de todos los principios mencionados tiene el mismo sentido: ellos permiten, de un modo natural, reducir la clase de las estimaciones sujetas a estudio, de manera que en las reducciones obtenidas resulte posible la determinación de las estimaciones eficientes.

1. **Estimaciones del parámetro de desplazamiento y escala.** Se llama problema de estimación del parámetro de desplazamiento el problema de estimación del parámetro α en la familia de distribuciones $\{\mathbf{P}_\alpha\}$ que poseen la propiedad

$$\mathbf{P}_\alpha(A) = \mathbf{P}(A - \alpha).$$

Aquí $\mathbf{P}$ es cierta distribución registrada; $A - \alpha = \{x : x + \alpha \in A\}$ y se supone que el conjunto paramétrico Θ tiene la misma naturaleza que $\mathscr{X}$. En el caso en que $\mathscr{X} = R^m$ se puede, por supuesto, examinar también los desplazamientos de θ de "menor dimensión", por ejemplo, escalares, pero entonces es necesario registrar la dirección (vector $e \in \mathscr{X}$) de desplazamiento y estudiar $\mathbf{P}_\alpha(A) = \mathbf{P}(A + \alpha e)$. Para abreviar, examinaremos tan sólo la primera posibilidad y consideraremos que $\Theta = \mathscr{X} = R^m$.

Señalemos que la distribución $\mathbf{P}_\alpha$ de $x_i + c (c \in R^m)$ coincide con la distribución $\mathbf{P}_{\alpha+c}$ de la magnitud x_i, o sea, el desplazamiento de todas las observaciones en c conduce a la muestra de la distribución $\mathbf{P}_{\alpha+c}$. Por eso es natural que se investiguen únicamente las estimaciones $\alpha^* = \alpha^*(X)$ del parámetro α que poseen la propiedad

$$\alpha^*(X + c) = \alpha^*(X) + c. \tag{1}$$

De aquí en adelante $X + c$ significará el vector con coordenadas $(x_1 + c, \ldots, x_n + c)$. La violación de esta igualdad significaría que la estimación

α^* depende del origen, o sea, de la elección del origen de coordenadas en el espacio $\mathscr{X} = R^n$.

El enfoque análogo aparece al estimar el parámetro de escala cuando se aprecia el parámetro σ en la familia $\{\mathbf{P}_\sigma\}$ que tiene la propiedad $\mathbf{P}_\sigma(A) = (A/\sigma)$, $\sigma \in (0, \infty)$. Aquí suponemos que σ es escalar, aunque se puede examinar también un caso matricial. En este caso la distribución $\mathbf{P}_\sigma$ de los valores $x_i c$ coincide con la distribución $\mathbf{P}_{\sigma c}$ de las magnitudes x_i, o sea, la multiplicación de las observaciones por c conduce a la muestra de $\mathbf{P}_{\sigma c}$. Por consiguiente, en este caso es natural limitarse al examen de las estimaciones que poseen la propiedad

$$\sigma^*(Xc) = c\sigma^*(X), \tag{2}$$

donde $Xc = (x_1 c, \ldots, x_n c)$, puesto que al variar c veces la escala de observaciones esa misma cantidad de veces también varía el parámetro.

El lector, por su propia iniciativa, puede obtener fácilmente las afirmaciones siguientes.

Si la familia $\mathbf{P}_\theta$ *satisface la condición* (A_μ), *entonces* θ *será de parámetro de desplazamiento (de escala) si y sólo si*

$$f_\theta(x) = f(x - \theta), \quad \left(f_\theta(x) = \frac{1}{\theta} f\left(\frac{x}{\theta}\right)\right).$$

Si $\mathscr{X} = R = \Theta$, $X \in \mathbf{P}_\alpha$ y α es el parámetro de desplazamiento, entonces $Y = e^X = (e^{x_1}, \ldots, e^{x_n}) \in \mathbf{Q}_\sigma$, donde, para las distribuciones $\mathbf{Q}_\sigma$, $\sigma = e^\alpha$ es el parámetro de escala. Esto se deduce directamente del hecho de que la densidad $y_1 = e^{x_1}$ es igual a (véase [11], p.)

$$\frac{1}{y} f(\ln - \alpha) = \frac{1}{\sigma}\left[\frac{\sigma}{y} f\left(\ln \frac{y}{\sigma}\right)\right].$$

Al contrario, si $\mathscr{X} = (0, \infty) = \Theta$, $X \in \mathbf{P}_\sigma$ y σ es el parámetro de escala, entonces $Y = \ln X = (\ln x_1, \ldots, \ln x_n) \in \mathbf{Q}_\alpha$, donde $\alpha = \ln \sigma$ es el parámetro de desplazamiento de las distribuciones $\mathbf{Q}_\alpha$.

Se puede examinar también el problema de estimación simultánea de los parámetros desconocidos α y σ en el caso en que $\mathbf{P}_{\alpha,\sigma}(A) = \mathbf{P}\left(\dfrac{A - \alpha}{\sigma}\right)$. En estas condiciones es natural que en calidad de estimación de σ se examinen las funciones que poseen la propiedad

$$\alpha^*(X + c) = \alpha^*(X), \quad \sigma^*(Xc) = c\sigma^*(X). \tag{3}$$

Las estimaciones que en los ejemplos examinados satisfacen las condiciones (1), (2) y (3) se llaman *equivariantes* (véase la definición general en el § 19). La causa de introducción de tales estimaciones consiste en la con-

tracción de todas las estimaciones sometidas a estudio, lo cual simplifica el problema de búsqueda de las estimaciones óptimas. Así en el § 8 hemos establecido que es imposible hallar uniformemente (o sea, para todos los θ) las mejores estimaciones en la clase de todas las estimaciones. Resulta que en la clase de estimaciones equivariantes tales estimaciones uniformemente mejores ya existen y en varios casos pueden ser halladas en forma explícita. Vamos a ilustrar este hecho citando, a título de ejemplo, las estimaciones de desplazamiento y escala.

2. Estimación eficiente del parámetro de desplazamiento en la clase de estimaciones equivariantes. Aquí consideraremos que se cumple la condición (A_μ) y, por lo tanto, $f_\alpha(x) = f(x - \alpha)$ y que μ es la medida de Lebesgue.

Designemos por S_0 la estadística

$$S_0 = S_0(X) = (x_2 - x_1, \ \ldots, \ x_n - x_1)$$

que es, evidentemente, invariante respecto al desplazamiento: $S_0(X + c) = S_0(X)$. Designemos por K_E la clase de todas las estimaciones equivariantes α^*, o sea, las estimaciones que satisfacen (1), y designemos por $|\alpha|^2$ el cuadrado de la norma euclídea $\alpha \in R^m$.

Teorema 1. *Sea* $\alpha^* = \alpha^*(X)$ *cualquier estimación equivariante con valor finito* $\mathbf{M}_0\alpha^*$. *Entonces, la estimación*

$$\alpha_0^* = \alpha^* - \mathbf{M}_0(\alpha^*/S_0) \tag{4}$$

no depende de la elección de α^* *y es la única estimación eficiente en la clase* K_E, *o sea,* $\mathbf{M}_\alpha|\alpha_0^* - \alpha|^2 = \min\limits_{\alpha^* \in K_E} \mathbf{M}_\alpha|\alpha^* - \alpha|^2$ *para todos los* α *y* $\mathbf{M}_\alpha|\alpha^* - \alpha|^2 = \mathbf{M}_\alpha|\alpha_0^* - \alpha|^2$ *si sólo* $\mathbf{M}_0(\dot{\alpha}^*/S_0) = 0$ *c.d. La estimación* α_0^* *puede ser representada en la forma*

$$\alpha_0^* = \frac{\int u f_u(X)du}{\int f_u(X)du} = \frac{\int u f(X - u)du}{\int f(X - u)du}. \tag{5}$$

La estimación α_0^* se denomina *estimación de Pitman.* De (4) es fácil deducir que ésta es equivariante y no está desplazada. La equivariación se deduce de la equivariación de α^* y de la invariación respecto al desplazamiento de la función $V(S_0) = \mathbf{M}_0(\alpha^*/S_0)$ que depende tan sólo de S_0. El no desplazamiento se deduce de las igualdades

$$\mathbf{M}_\alpha\alpha_0^* = \alpha + \mathbf{M}_\alpha\alpha^*(X - \alpha) - \mathbf{M}_\alpha V(S_0), \tag{6}$$

donde $\mathbf{M}_\alpha V(S_0) = \mathbf{M}_0 V(S_0)$, $\mathbf{M}_\alpha\alpha^*(X - \alpha) = \mathbf{M}_0\alpha^*(X)$. La última relación se deduce del hecho de que $X - \alpha \in \mathbf{P}_0$ si $X \in \mathbf{P}_\alpha$. Por eso la suma de los dos últimos sumandos en (6) constituye

$$\mathbf{M}_0\alpha^* - \mathbf{M}_0[\mathbf{M}_0(\alpha^*/S_0)] = 0; \ \mathbf{M}_\alpha\alpha_0^* = \alpha.$$

Antes de demostrar el teorema expondremos la siguiente afirmación auxiliar.

Lema 1. *Sea $X \in \mathbf{P}_0$. Para cualquier estadística $S = S(X)$ con esperanza matemática finita $\mathbf{M}_0|S| < \infty$, la e.m.c. de S respecto a S_0 es igual a*

$$\mathbf{M}_0(S/S_0) = S_1(X) \equiv \frac{\int S(X-u)f_u(X)du}{\int f_u(X)du}. \tag{7}$$

Demostración. Todas las funciones bajo los signos integrales en (7) son las funciones de $X - u$. Por consiguiente, después de sustituir $x_1 - u = v$, las mismas serán las funciones de $(v, x_2 - x_1 + v, \ldots, x_n - x_1 + v)$. Esto quiere decir que el segundo miembro de (7) depende únicamente de S_0. En virtud de las propiedades de la e.m.c., para demostrar el lema es suficiente convencerse que para cualquier $A \in \sigma(S_0)$

$$\mathbf{M}_0(S_1;\, A) = \mathbf{M}_0(S;\, A). \tag{8}$$

Sea $Z = Z(S_0)$ cualquier estadística $\sigma(S_0)$-medible limitada. Entonces

$$\mathbf{M}_0 Z S_1 = \int_{\mathscr{X}^n} \frac{Z(S_0)\int_\Theta S(x-u)f_u(x)du}{\int_\Theta f_u(x)du} f(x)\,dx =$$

$$= \int_\Theta \int_{\mathscr{X}^n} \frac{Z(S_0)S(x-u)f(x-u)f(x)}{\int_\Theta f(x-v)dv}\,dx\,du.$$

Después de sustituir $x - u \to x$, en el intervalo interior obtenemos (en este caso $S_0(x)$ se transforma en sí mismo)

$$\int_\Theta \int_{\mathscr{X}^n} \frac{Z(S_0)S(x)f(x)f(x+u)}{\int_\Theta f(x+u-v)dv}\,dx\,du = \int_{\mathscr{X}^n} Z(S_0)S(x)f(x)dx = \mathbf{M}_0 ZS.$$

Esto demuestra (8). El cambio del orden de integración, al cual hemos acudido dos veces, es justo en virtud de la integrabilidad absoluta de S y del carácter limitado de Z. ◁

Demostración del teorema 1. Antes que nada es preciso señalar que para la estimación equivariante, $\mathbf{M}_\alpha|\alpha^* - \alpha|^2$ no depende de α. En efecto,

$$\mathbf{M}_\alpha|\alpha^*(X) - \alpha|^2 = \mathbf{M}_\alpha|\alpha^*(X-\alpha)^2 = \mathbf{M}_0|\alpha^*(X)|^2.$$

Ahora bien, para determinar la estimación equivariante uniformemente óptima es necesario hallar α^*, que minimiza $\mathbf{M}_0|\alpha^*|^2$.

Sea α^* cualquier estimación equivariante α. En virtud de las propiedades de la e.m.c.,

$$\mathrm{M}_0|\alpha^*|^2 = \mathrm{M}_0|\alpha^* - \mathrm{M}_0(\alpha^*/S_0)|^2 + \mathrm{M}_0|\mathrm{M}_0(\alpha^*/S_0)|^2 \geqslant$$
$$\geqslant \mathrm{M}_0|\alpha^* - \mathrm{M}_0(\alpha^*/S_0)|^2. \quad (9)$$

Queda señalar que, en virtud del lema 1, la estimación $\alpha_0^* = \alpha^* - \mathrm{M}_0(\alpha^*/S_0)$ es igual a (5) y no depende de la elección de α^*. La igualdad en (9) es, evidentemente, posible si y sólo si $\mathrm{M}_0(\alpha^*/S_0) = 0$ c.d. ◁

De la demostración del teorema se deduce que, en la construcción de la estimación óptima equivariante, desempeña un papel especial la estadística $S_0 = (x_2 - x_1, \ldots, x_n - x_1)$, que es invariante respecto a la transformación del desplazamiento. La invariación de la estadística es, en cierto sentido, una cualidad contraria a la suficiencia, y la construcción de la estimación $\theta_0^* = \theta^* - \mathrm{M}_0(\theta^*/S_0)$ a base de la estimación arbitraria θ^*, es el enfoque del mejoramiento de la estimación θ^*, también, en cierto sentido, contrario al enfoque con el cual, para el mejoramiento de la estimación θ^* mediante la estadística suficiente S, se examina la estimación $\theta_s^* = \mathrm{M}_\theta(\theta^*/S)$. La contrariedad consiste en lo siguiente. La característica suficiente contiene toda la información sobre el parámetro θ, mientras que la estadística invariante no contiene ninguna. Con el fin de obtener las mejores estimaciones, hemos buscado las estadísticas suficientes mínimas; aquí, como veremos, necesitamos las estadísticas invariantes máximas (tal es la estadística S_0). La estimación θ_s^* es la «proyección» de θ^* sobre S, mientras que la estimación θ_0^* se obtiene sustrayendo de θ^* su «proyección» sobre S_0.

En resumidas cuentas, los resultados obtenidos por estas dos vías coinciden a menudo, como se verá de los dos ejemplos siguientes.

Ejemplo 1. Sea $\mathscr{X} = R$, $X \in \Phi_{\alpha,1}$. Entonces

$$f_\alpha(X) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum (x_i - \alpha)^2\right\} =$$
$$= \frac{1}{\sqrt{n}\,(2\pi)^{\frac{n-1}{2}}} \exp\left\{-\frac{1}{2}\sum (x_i - \bar{x})^2\right\} \cdot \sqrt{\frac{n}{2\pi}}\, e^{-\frac{n}{2}(\alpha-\bar{x})^2}.$$

Aquí el segundo factor, como función de α, es la función de densidad de la ley normal con parámetros $(\bar{x}, 1/n)$. Como el primer factor no depende de α, es reducido en (5), y la estimación de Pitman constituirá $\alpha^* = \bar{x}$. En el caso multidimensional obtendremos este mismo resultado.

Ejemplo 2. Sea $\mathscr{X} = R$, $X \in \mathrm{U}_{\theta,1/\theta}$. Entonces

$$f_\theta(X) = \begin{cases} 1 & \text{cuando } x_{(n)} - 1 \leqslant \theta \leqslant x_{(1)}, \\ 0 & \text{en los demás casos.} \end{cases}$$

Por eso

$$\theta^* = \int_{x_{(n)}-1}^{x_{(1)}} u\,du \Big/ (x_{(1)} - x_{(n)} + 1) = \frac{1}{2}(x_{(1)} + x_{(n)} - 1).$$

Ahora bien, vemos que en la clase K_E de estimaciones equivariantes se pueden construir, en forma explícita, las estimaciones eficientes, además, en este caso no se necesitan ningunas condiciones de suavidad de $f_\theta(x)$, y la propia eficacia tiene un carácter exacto (no asintótico).

3. Carácter minimax de la estimación de Pitman. Ahora prestemos atención a la forma de estimación de Pitman. Hablando en términos generales, ésta es una estimación bayesiana para la distribución a priori «uniforme en todo el eje». Como tal distribución no existe, enunciemos más exactamente la referida afirmación. Sea $\mathscr{X} = R$ y $\mathbf{Q}^{(N)}$ una distribución uniforme en $[-N, N]$, o sea, una distribución cuya densidad constituye

$$q^{(N)}(t) = \begin{cases} (2N)^{-1}, & |t| \leqslant N, \\ 0 & |t| > N. \end{cases}$$

La estimación bayesiana correspondiente a $\mathbf{Q}^{(N)}$ será igual a

$$\alpha^*_{Q^{(N)}} = \frac{\int u q^{(N)}(u) f_u(X) du}{\int q^{(N)}(u) f_u(X) du} = \int_{-N}^{N} u f_u(X) du \Big/ \int_{-N}^{N} f_u(X) du.$$

Es evidente que para todos X, la estimación de Pitman α_0^* es el límite $\alpha_0^* = \lim_{N\to\infty} \alpha^*_{Q^{(N)}}$. Esta circunstancia sugiere que a la vez convergerán también los momentos de segundo orden:

$$\mathbf{M}_\alpha(\alpha^*_{Q^{(N)}} - \alpha)^2 \to \mathbf{M}_\alpha(\alpha_0^* - \alpha)^2.$$

Resulta que en la región $|\alpha| \leqslant N - \sqrt{N}$, eso es precisamente así. Además, la convergencia será uniforme respecto a α en el referido intervalo de valores de α. (La demostración está relacionada con la estimación de $\mathbf{M}_\alpha(\alpha_0^* - \alpha^*_{Q^{(N)}})^2$, tiene principalmente carácter técnico y por eso la omitimos).

Pero en este caso podemos utilizar el criterio del carácter minimax de las estimaciones en el teorema 11.3; *si la estimación α^* es tal que, para todos los valores de α,*

$$\mathbf{M}_\alpha(\alpha^* - \alpha)^2 \leqslant \lim_{N\to\infty} \sup \int \mathbf{M}_t(\alpha^*_{Q^{(N)}} - t)^2 \mathbf{Q}^{(N)}(dt) \qquad (10)$$

para cierta sucesión de distribuciones a priori $\mathbf{Q}^{(N)}$ (no obligatoriamente uniformes) y de estimaciones bayesianas correspondientes $\alpha^*_{Q^{(N)}}$, entonces *α^* es una estimación minimax.*

En nuestro caso, $m = \mathbf{M}_\alpha(\alpha_0^* - \alpha)^2$ no depende de α. Por eso, en virtud de las propiedades de convergencia anteriormente mencionadas,

$$\lim_{N\to\infty} \sup \int \mathbf{M}_t(\alpha^*_{Q^{(N)}} - t)^2 \mathbf{Q}^{(N)} dt) \geqslant$$

$$\geqslant \lim_{N\to\infty} \sup \frac{1}{2N} \int_{|t|<N-\sqrt{N}} \mathbf{M}_t(\alpha^*_{Q^{(N)}} - t)^2 dt \geqslant$$

$$\geqslant \lim_{N\to\infty} \sup \frac{1}{2N} 2(N - \sqrt{N})(m - \varepsilon) = m - \varepsilon$$

para cualquier $\varepsilon > 0$. Esto significa que se ha cumplido la propiedad (10).

Así pues, la *estimación de Pitman es minimax en la clase de todas las estimaciones del parámetro de desplazamiento* (el hecho de que ella sea minimax en la clase de estimaciones equivariantes, se desprende, evidentemente, de la eficacia).

Lo dicho también se puede interpretar del modo siguiente: la «peor» distribución a priori (véase el § 11) para el parámetro de desplazamiento es la distribución «uniforme en todo el eje».

Como indicación del carácter minimax de la estimación de Pitman también podría servir la dependencia (señalada más arriba) $\mathbf{M}_\alpha(\alpha_0^* - \alpha)^2$ de α (compárese con el teorema 11.2).

4. Acerca de las estimaciones óptimas del parámetro de escala. Como ya hemos indicado, el problema de estimación del parámetro de escala σ puede reducirse, en cierto sentido, al problema de estimación del parámetro de desplazamiento. Sea, por abreviar, $\mathscr{X} = (0, \infty) = \Theta$. En este caso, si $X \in \mathbf{P}_\sigma$, $\mathbf{P}_\sigma(A) = \mathbf{P}(A/\sigma)$, entonces $Y = \ln X = (\ln x_1, \ldots, \ln x_n) \in \mathbf{P}_\alpha^{(1)}$, donde $\alpha = \ln \sigma$, y la distribución $\mathbf{P}_\alpha^{(1)}$ tiene una densidad $y_1 = \ln x_1$ en el punto y $\left(\text{la condición } (A_\mu) \text{ se cumple, } \frac{d\mathbf{P}_1(x)}{d\mu} = f(x)\right)$, igual a (véase [11], pág. 53)

$$f\left(\frac{e^y}{\sigma}\right)\frac{e^y}{\sigma} = f(e^{y-\alpha})e^{y-\alpha} = f^{(1)}(y - \alpha),$$
$$f^{(1)}(y) = f(e^y)e^y.$$

Ahora bien, podemos apreciar muy bien el parámetro α con ayuda de la estimación de Pitman $\alpha^* = \alpha^*(Y)$, y luego suponer que $\sigma^*(X) = e^{\alpha^*(Y)}$. Es fácil notar que $\sigma^*(X)$ será equivariante, ya que

$$\sigma^*(cX) = e^{\alpha^*(Y+\ln c)} = e^{\alpha^*(Y)+\ln c} = c\sigma^*(X).$$

No obstante, aquí es importante señalar que la estimación de Pitman mini-

miza $M_\alpha(\alpha^* - \alpha)^2$. Por lo tanto, la estimación σ^* obtenida minimizará

$$M_\sigma\left(\ln \frac{\sigma^*}{\sigma}\right)^2 \tag{11}$$

y no la magnitud $M_\sigma(\sigma^* - \sigma)^2$ de la cual se trataba generalmente. Pero en el problema de estimación equivariante del parámetro σ no era racional examinar la estimación estándar, puesto que ella, a distinción de (11), depende de la transformación de contracción aplicada simultáneamente a σ^* y σ. Aquí, como análogo de la estadística invariante S_0 servirá la estadística $(x_2/x_1, \ldots, x_n/x_1)$. A la par con (11) también es posible, naturalmente, examinar otros errores. Si, por ejemplo, minimizamos la magnitud

$$M_\sigma\left(\frac{\sigma^*}{\sigma} - 1\right)^2,$$

entonces, la mejor estimación equivariante será

$$\sigma^* = \frac{\int \sigma^{-n-2} f(X/\sigma)d\sigma}{\int \sigma^{-n-3} f(X/\sigma)d\sigma} \tag{12}$$

(véase [33], p.   ).

Ejemplo 3. Detección de la fuente de radiación. Examinemos un ejemplo de un problema físico real, relacionado con las estimaciones de desplazamiento y escala.

Supongamos que en cierto punto desconocido z del espacio tridimensional se encuentra una fuente de radiación gamma. El problema consiste en determinar las coordenadas del punto z utilizando un detector plano (que coincide con uno de los planos de coordenada) y, fijando en este detector las trazas de radiación, o sea, las trazas de interacción de los cuantos gamma, emitidos por el punto z, con la superficie sensible del detector.

Este problema sería mucho más simple si tuviéramos una fuente de radiación de partículas cargadas de alta energía. Entonces podríamos poner, uno tras otro, dos detectores planos paralelos y fijar en ellos los puntos de paso (o sea, de interacción con la superficie de la pantalla) tan sólo de dos partículas. Esto nos daría las direcciones del vuelo de esas partículas y junto con ellas las coordenadas del punto z como punto de intersección de dichas direcciones. Sin embargo, para una radiación gamma poco intensa, que se utiliza en roentgenoscopia, esto es irrealizable y tan sólo se puede introducir un detector.

La dirección de propagación de los cuantos gamma emitidos es aleatoria y se distribuye uniformemente en la superficie de la esfera (si dicha dirección se determina por un punto en la esfera con centro en el punto z).

Para simplificar el problema examinemos su variante bidimensional. Supongamos que la fuente se encuentra en el plano de las variables (x, y),

en un punto desconocido $z = (\alpha, \sigma)$, $\sigma > 0$. El ángulo de dirección de la radiación, formado con el eje Oy, tiene una distribución uniforme en $[0, 2\pi]$. El detector sensible coincide con el eje de abscisas. Los resultados de las observaciones serán los puntos $x_1, x_2, \ldots$, en los que hemos fijado la interacción de los cuantos gamma con el detector (con el eje de abscisas).

La peculiaridad de este problema consiste en que el volumen n de la muestra obtenida durante un tiempo fijo t, será aleatorio: el número de cuantos gamma emitidos por la fuente en el tiempo t tiene una distribución de Poisson, y el número de cuantos gamma que alcanzaron el detector también está distribuido con arreglo a la ley de Poisson, ya que cada cuanto llega al eje de abscisas con una probabilidad igual a 1/2. No obstante, en nuestro caso, n y las observaciones $x_1, x_2, \ldots$ son independientes. Por eso podemos examinar el número n de observaciones que se ha obtenido y considerarlo fijo (para cada uno de tales números n fijos, la distribución de x_i será la misma).

Así pues, supongamos que se han dado las observaciones $X = (x_1, \ldots, x_n)$. Nuestro problema consiste en estimar las coordenadas (α, σ). Mostremos que $X \in K_{\alpha,\sigma}$, o sea, x_i tienen una distribución de Cauchy con parámetros de desplazamiento α y de escala σ.

En efecto, la distribución condicional del ángulo β entre la dirección del movimiento del cuanto gamma y el eje $(0, -y)$, a condición de que

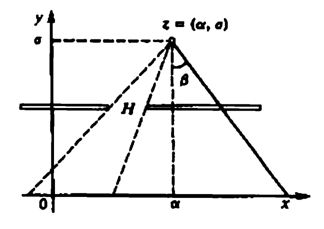


Fig. 2.

el cuanto haya alcanzado el detector (el eje de abscisas), será uniforme en el segmento $[-\pi/2, \pi/2]$. Como $(x - \alpha)/\sigma = \operatorname{tg} \beta$ (véase la fig. 2), entonces

$$\mathbf{P}_{\alpha,\sigma}(x_1 < x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} \frac{x - \alpha}{\sigma}.$$

Por consiguiente, la densidad de distribución de x_1 será igual a la densidad de distribución de Cauchy (véase el § 2)

$$k_{\alpha,\sigma}(x) = \frac{1}{\pi\sigma} \frac{1}{(1 + ((x - \alpha)/\sigma)^2)} = \frac{\sigma}{\pi(\sigma^2 + (x - \alpha)^2)}.$$

Ahora supongamos que σ es conocido, por ejemplo, $\sigma = 1$. Entonces la mejor estimación invariante del parámetro de desplazamiento α será la de Pitman, que se obtiene como el valor medio de $\alpha^* = \int u\varphi(u)du$ de la distribución con una densidad de

$$\varphi(u) = \varphi(u, X) = \frac{k_u(X)}{\int k_v(X)dv}, \quad k_u(X) = \prod_{i=1}^{n} k_u(x_i),$$
$$k_u(x_i) = k_{u,1}(x_i) = \frac{1}{\pi(1 + (u - x_i)^2)}.$$

La e.v.m $\hat{\alpha}^*$ será un punto en el que se alcanza el máx $\varphi(u)$. Más adelante mostraremos (véanse los §§ 24 y 25) que α^* y $\hat{\alpha}^*$ son asintóticamente equivalentes y tienen una distribución asintóticamente normal con coeficiente $1/I = 2$ (en el caso sujeto a examen $I = \int (k_0')^2/k_0 dx = 4\pi^{-1}\int x^2(1 + x^2)^{-3}dx = 1/2$). De lo dicho resulta que el error de las estimaciones α^* y $\hat{\alpha}^*$ para grandes n tiene un orden de pequeñez igual a $1/\sqrt{n}$.

Es interesante señalar que en el problema sometido a examen se puede alcanzar un grado más alto de exactitud, interviniendo en el experimento. Esto se puede hacer colocando entre el punto $z = (\alpha, 1)$ y el detector una pantalla paralela al eje de abscisas y provista del orificio H, a través del cual sólo pueden pasar los cuantos gamma. Las posiciones de la pantalla y el orificio se eligen según el experimentador y, por lo tanto, son conocidas.

En este caso la distribución de las observaciones en la pantalla será discontinua y, si los orificios H son pequeños, será próxima a $U_{a\alpha, a\alpha+b}$ para ciertas constantes a y b que conocemos. La forma de la estimación equivariante eficiente α_H^* para tal distribución fue hallado en el ejemplo 2. La estimación α_H^* se determina por los valores extremos de la muestra y tiene una exactitud del orden de $1/n_H$, donde $n_H \leqslant n$ es el número de elementos de la muestra, los cuales corresponden a los cuantos que han pasado a través de la ranura (n_H, al igual que n, es realmente aleatorio y está distribuido de acuerdo con la ley de Poisson). Como, por término medio, n_H es proporcional a n, con valores de n bastante grandes obtenemos $1/n_H \ll 1/\sqrt{n}$.

§ 19*. Problema general sobre la estimación equivariante

Examinemos el grupo G de transformaciones medibles g del espacio $\mathscr{X}^n$ en sí, que poseen las propiedades siguientes:

1) cada g aplica $\mathscr{X}^n$ en todo el espacio $\mathscr{X}^n$, o sea, para cada $x_2 \in \mathscr{X}^n$ se encontrará un $x_1 \in \mathscr{X}^n$ tal que $x_2 = gx_1$.

2) las aplicaciones g son biunívocas.

La mensurabilidad de g se necesita para que gX sea una variable aleatoria. La propiedad de grupo quiere decir que $g_2g_1 \in G$ si $g_1 \in G$, $g_2 \in G$; la transformación idéntica e y la inversa g^{-1} pertenecen a G (así que $g^{-1}g = e$).

Definición 1. La familia de distribuciones $\{\mathbf{P}_\theta\}$ se llama *equivariante respecto al grupo de transformaciones* G(o, para abreviar, simplemente invariante) si para cada $g \in G$ y $\theta \in \Theta$ existe el único $\theta_g \in \Theta$ tal que la relación $X \Subset \mathbf{P}_\theta$ conduce a $gX \Subset \mathbf{P}_{\theta_g}$.

Designemos por $\theta_g = \bar{g}\theta$ el valor de θ_g definible unívocamente por θ y g. Entonces la definición significa que

$$\mathbf{P}_\theta(gX \in A) = \mathbf{P}_{\bar{g}\theta}(X \in A).$$

Como en virtud de la definición 1 se cumple la condición (A_0), el conjunto $\bar{G}$ de todas las transformaciones $\bar{g}$ del espacio Θ en sí forma un grupo. En efecto, la distribución g_2g_1X se da simultáneamente por las distribuciones $\mathbf{P}_{\overline{g_2g_1}\theta}$ y $\mathbf{P}_{\bar{g}_2\bar{g}_1\theta}$. De la condición ($A_0$) resulta que $\overline{g_2g_1} = \bar{g}_2\bar{g}_1$ y que $\bar{g}_1^{-1} \in \bar{G}$ (es suficiente poner $g_2 = g_1^{-1}$). Las transformaciones $\bar{g}$ de $\bar{G}$ son automáticamente biunívocas. Sin embargo, puede no haber isomorfismo entre G y $\bar{G}$. Sea, por ejemplo, $X \Subset \Phi_{0,\sigma^2}$, $\sigma \in (0, \infty)$. En este caso la densidad $f_{0,\sigma^2}(X)$ (función de verosimilitud) depende exclusivamente de $\sum x_i^2$. Por consiguiente, si en calidad de G examinamos un grupo de revoluciones (transformaciones ortogonales de $\mathscr{X}^n$), entonces, las condiciones de la definición 1 serán cumplidas, pero $\bar{g} = \bar{e}$, y el grupo $\bar{G}$ se compone del único elemento $\bar{e}$, o sea, de la transformación idéntica de $\Theta = (0, \infty)$ en sí.

Le proponemos al lector que compruebe, en calidad de ejercicio, que si $\{\mathbf{P}_\theta\}$ es invariante respecto al grupo G, y G_1 es un subgrupo de G, entonces $\{\mathbf{P}_\theta\}$ es invariante respecto a G_1.

Cuando examinemos el problema general de estimación equivariante necesitaremos un planteamiento más general del problema respecto a la comparación de las estimaciones. Hasta ahora lo hemos hecho con ayuda de las desviaciones estándar, midiendo el error de la estimación por la magnitud $(\theta^* - \theta)^2$. Ahora supondremos que la medición del error de θ^* ocurre con ayuda de la función $w(\theta^*, \theta)$ y que esta función posee propiedad de "homogeneidad"*):

$$w(\bar{g}\theta, \bar{g}\theta^*) = w(\theta, \theta^*) \text{ para todos los valores de } \theta. \qquad (1)$$

Precisamente esta propiedad es típica de las funciones $w(\theta, \theta^*) = (\theta - \theta^*)^2$ para el parámetro de desplazamiento (transformación de desplazamiento)

---

*) Esta propiedad no es obligatoria en la teoría de estimación equivariante. Sólo se puede exigir la existencia de $\bar{g}\theta^*$ tal que para todos θ $w(\bar{g}\theta, \bar{g}\theta^*) = w(\theta, \theta^*)$ (véase [33]).

y $w(\theta, \theta^*) = \left(\ln \frac{\theta}{\theta^*}\right)^2$ ó $\left(\frac{\theta}{\theta^*} - 1\right)^2$ para el parámetro de escala (transformación de contracción).

Hemos visto en el punto 4 del § 18, que el problema de determinación de la mejor estimación invariante puede ser muy sensible al elegir la medida del error $w(\theta, \theta^*)$ de la estimación θ^*.

Recurramos ahora al problema de estimación de las familias invariantes $\{\mathbf{P}_\theta\}$. Supongamos que tenemos la muestra X y que basándonos en ella hemos construido la estimación $\theta^* = \theta^*(X)$ del parámetro θ. Si examinamos la muestra $Y = gX \in \mathbf{P}_{\bar{g}\theta}$, entonces $\theta^*(Y)$ será la estimación para $\bar{g}\theta$. En este caso es natural suponer que las estimaciones $\theta^*(X)$ y $\theta^*(Y)$ están ligadas entre sí al igual que los parámetros sujetos a estimación θ y $\bar{g}\theta$, o sea, mediante la transformación $\bar{g}$:

$$\theta^*(Y) = \bar{g}\theta^*(X). \tag{2}$$

En virtud de (1), la estimación $\theta^*(Y)$ del parámetro $\bar{g}(\theta)$ proporciona el mismo error que la estimación $\theta^*(X)$ del parámetro θ. Por lo tanto, tenemos dos problemas de estimación "iguales". Las transformaciones realizadas gX y $\bar{g}\theta$ pueden interpretarse como las sustituciones de los sistemas de coordenadas. Entonces (2) significa que la estimación θ^* no depende de la elección del sistema de coordenadas y satisface la relación

$$\theta^*(X) = \bar{g}^{-1}\theta^*(gX). \tag{3}$$

Con otras palabras, si se ha elegido θ^*, que satisface (2), entonces no importa cuál de los dos problemas de estimación mencionados más arriba ha de ser resuelto, puesto que, mediante la igualdad (3), las deducciones acerca de $\bar{g}\theta$ en el segundo problema pueden convertirse en deducciones acerca de θ en el primer problema.

Definición 2. La estimación θ^* del parámetro θ de la familia invariante $\mathbf{P}_\theta$ que satisface (3) se llama *equivariante**⁾.

Examinemos cualquier punto $\theta_0 \in \Theta$ y el conjunto de puntos "equivalentes" $\theta = \bar{g}\theta_0$, $\bar{g} \in \overline{G}$. Tal formación de clases de puntos "equivalentes" divide todo el espacio Θ en subconjuntos llamados *órbitas*.

Teorema 1. *El valor de* $\mathbf{M}_\theta w(\theta, \theta^*)$ *para la estimación equivariante* θ^* *es constante en la órbita, o sea,*

$$\mathbf{M}_\theta w(\theta, \theta^*) = \mathbf{M}_{\bar{g}\theta} w(\bar{g}\theta, \theta^*)$$

para cualesquiera $\theta \in \Theta$ *y* $\bar{g} \in \overline{G}$.

*⁾ Tales estimaciones se denominan, a veces, invariantes. Sin embargo, este término es menos exacto. Es mejor dejarlo para las estimaciones que poseen la propiedad $\theta^*(gX) = \theta^*(X)$ (o sea, para el caso cuando $\bar{g} = \bar{e}$ para todo g).

Demostración.

$$\mathbf{M}_\theta w(\theta,\ \theta^*(X)) = \mathbf{M}_\theta w(\bar{g}\theta,\ \bar{g}\theta^*(X)) =$$
$$= \mathbf{M}_\theta w(\bar{g}\theta,\ \theta^*(gX)) = \mathbf{M}_{\bar{g}\theta} w(\bar{g}\theta,\ \theta^*(X)). \quad \triangleleft$$

Si la órbita $\{\theta;\ \theta = \bar{g}\theta_0,\ \bar{g} \in \overline{G}\}$ coincide con Θ (como tuvo lugar para los parámetros de desplazamiento y escala), entonces $\mathbf{M}_\theta w(\theta,\ \theta^*) =$ const en Θ. El cumplimiento de esta igualdad es el síntoma característico del carácter minimax de θ^* (compárese con el teorema 11.2), así que las mejores estimaciones equivariantes a menudo resultan minimax en la clase de todas las estimaciones (esto se detalla en [33]).

De los teoremas del § 11 se deduce, por ejemplo, el

Teorema 2. *Si Θ es una órbita, y la estimación equivariante θ^* resultó bayesiana (o el límite de estimaciones bayesianas θ_N^* con una convergencia $\mathbf{M}_\theta w(\theta,\ \theta^*) = \lim_{N\to\infty} \mathbf{M}_\theta w(\theta,\ \theta_N^*)$), entonces θ^* es una estimación minimax.*

Nótese también la siguiente propiedad importante de las estimaciones equivariantes. Será cómodo designar por $\nu(g,\ dx)/\nu(dx)$ la densidad de la medida ν_g, $\nu_g(B) = \nu(gB)$ respecto a la medida ν en el punto $x \in \mathscr{X}^n$.

Teorema 3. *Supongamos que se cumple la condición (A_μ) y $\mu^n(g\ dx)/\mu^n(dx)$ es finito y positivo para cada $g \in G$, y c.t. $[\mu^n]$ valores de x. Supongamos, además, que la e.v.m. $\hat{\theta}^*$ es la única para cada X. En este caso, si la familia $\mathbf{P}_\theta$ es invariante, entonces $\hat{\theta}^*$ es la estimación equivariante.*

Demostración. Tenemos

$$f_{\hat{\theta}^*}(X) = \frac{\mathbf{P}_{\hat{\theta}^*(X)}(dx)}{\mu^n(dx)} = \max_\theta \frac{\mathbf{P}_\theta(dx)}{\mu^n(dx)} \tag{4}$$

en el punto $x = X$. Suponiendo $Y = gX$, también podemos escribir

$$f_{\hat{\theta}^*(Y)}(Y) = \frac{\mathbf{P}_{\hat{\theta}^*(Y)}(g\ dx)}{\mu^n(g\ dx)} = \max_\theta \frac{\mathbf{P}_\theta(g\ dx)}{\mu^n(g\ dx)}.$$

En virtud de la invariación de $\mathbf{P}_\theta$ y del carácter finito de $\mu^n(g\ dx)/\mu^n(dx) > 0$, esto equivale a que

$$\frac{\mathbf{P}_{\bar{g}^{-1}\hat{\theta}(Y)}(dx)}{\mu^n(dx)} = \max_\theta \frac{\mathbf{P}_{\bar{g}^{-1}_{\theta}}(dx)}{\mu^n(dx)} = \max_\theta \frac{\mathbf{P}_\theta(dx)}{\mu^n(dx)}.$$

Comparando con (4) y utilizando la unicidad de $\hat{\theta}^*(X)$, obtenemos $\bar{g}^{-1}\hat{\theta}^*(gX) = \hat{\theta}^*(X)$. $\triangleleft$

§ 20. Desigualdad integral del tipo Rao—Cramer. Criterios del carácter asintóticamente bayesiano y minimax de las estimaciones

Este párrafo también podría titularse "Desigualdad para la desviación estándar en el caso bayesiano". En su mayor parte el mismo se refiere a la teoría asintótica de la estimación.

Antes ya hemos tocado las cuestiones relacionadas con el enfoque asintótico de la comparación de las estimaciones. Ahora, y sobre todo en los §§ 23—29, dichas cuestiones serán el principal objeto de estudio.

1. Estimaciones eficientes y superericientes. En el § 16, dedicado a la desigualdad de Rao—Cramer, quedó sin aclarar la siguiente cuestión importante. Supongamos que se cumple la cuestión (R). Entonces, para las estimaciones no desplazadas,

$$\mathbf{M}_\theta(\theta^* - \theta)^2 \geqslant \frac{1}{nI(\theta)}.$$

El segundo miembro de dicha desigualdad se llama, a veces, *frontera de Rao—Cramer.* Esta se alcanza para las estimaciones R-eficientes. La cuestión consiste en si ¿será posible o no, a costa de elegir el desplazamiento, mejorar considerablemente las estimaciones R-eficientes o asintóticamente R-eficientes? Es la cuestión acerca del carácter esencial de la frontera de Rao—Cramer y acerca del papel que desempeña el desplazamiento.

Ya hemos examinado parcialmente el hecho de que en un punto registrado θ_0, el valor de $\mathbf{M}_\theta(\theta^* - \theta)^2$ puede hacerse mucho menor que la frontera de Rao—Cramer. Para ello es suficiente tomar $\theta^* = \theta_0$. No obstante, en este caso, tal estimación en otros puntos será muy mala.

Se puede citar otro ejemplo menos trivial, donde el mejoramiento se alcanza no a expensas de otros puntos. Sea $X \in \Phi_{\alpha,1}$, $\alpha \in \Theta = [0, \infty)$. Entonces la estimación $\alpha^* = \bar{x}$ es eficiente e incluso R-eficiente. Sin embargo, en nuestro caso, cuando $\Theta = [0, \infty)$, la estimación $\alpha^{**} = \text{máx}(0, \bar{x})$ será, evidentemente, mejor, puesto que ella reduce las desviaciones estándar, sustituyendo por 0 los valores negativos inadmisibles. Es evidente que la estimación α^{**} ya será desplazada: $\mathbf{M}_\alpha\alpha^{**} > \alpha$, pero en el punto $\alpha = 0$ tenemos $I(\alpha) = 1$, $\mathbf{M}_0(\alpha^*)^2 = \frac{1}{n}$, $\mathbf{M}_0(\alpha^{**})^2 = \frac{1}{2n} < \frac{1}{nI(0)}$. En este ejemplo, el mejoramiento está relacionado con el hecho de que hemos reducido el campo de valores de la estimación α^* hasta el conjunto Θ. Citemos un ejemplo más (perteneciente a Hodges), en el que el mejoramiento de α^* ocurre no a costa de la limitación de Θ.

Sea, como antes, $X \in \Phi_{\alpha,1}$, $\alpha \in \Theta = (-\infty, \infty)$. Además de la estimación eficiente $\alpha^* = \bar{x}$ examinemos, cuando $\beta < 1$, la estimación

$$\alpha^{**} = \begin{cases} \bar{x} \text{ si } |\bar{x}| \geqslant n^{-1/4}, \\ \beta\bar{x} \text{ si } |\bar{x}| < n^{-1/4}. \end{cases}$$

No es difícil ver que, cuando $\alpha > 0$, según el teorema central, del límite,

$$\mathbf{P}_\alpha(|\bar{x}| < n^{-1/4}) \leqslant \mathbf{P}_\alpha((\bar{x} - \alpha)\sqrt{n} < n^{1/4} - \alpha\sqrt{n}) \to 0$$

cuando $n \to \infty$. La afirmación análoga es cierta cuando $\alpha < 0$. Por eso α^{**}, cuando $\alpha \neq 0$ α^{**}, coincide con $\bar{x}$ en el conjunto de la probabilidad que converge hacia 1 y, por lo tanto, según el teorema de continuidad cuando $\alpha \neq 0$,

$$(\alpha^{**} - \alpha)\sqrt{n} \in \Phi_{0,1}.$$

Cuando $\alpha = 0$,

$$\mathbf{P}_0(|\bar{x}| < n^{-1/4}) = \mathbf{P}_0(|\bar{x}\sqrt{n}| < n^{1/4}) \to 1$$

y α^{**} en el conjunto de la probabilidad convergente hacia 1 coincide con $\beta\bar{x}$, así que $(\alpha^{**} - \alpha)\sqrt{n} \in \Phi_{0,\beta^2}$. Por consiguiente, para todos los valores de α, la estimación α^{**} es asintóticamente normal, $(\alpha^{**} - \alpha)\sqrt{n} \in$ $\in \Phi_{0,\sigma^2(\alpha)}$, donde

$$\sigma^2(\alpha) = \begin{cases} 1 & \text{cuando } \alpha \neq 0, \\ \beta^2 \quad 1 & \text{cuando } \alpha = 0. \end{cases}$$

Ahora bien, en el punto $\alpha = 0$, el coeficiente de dispersión $\sigma^2(0)$ resultó menor que la frontera inferior de Rao—Cramer, igual a 1.

Las estimaciones asintóticamente normales en los ejemplos citados, cuando el coeficiente de dispersión para ellas $\sigma^2(\theta) \leqslant I^{-1}(\theta)$ es, con algunos valores de θ, estrictamente menor que $I^{-1}(\theta)$, se llaman, a veces, *superficientes*.

No obstante, resultó que estos ejemplos cambian poco el cuadro, justo en general, acerca de la preferencia de las estimaciones eficientes. Precisamente Le Cam demostró que el mejoramiento (ilustrado más arriba) de las estimaciones, hablando en general, sólo se puede lograr en pequeñas cantidades de puntos.

En este párrafo mostraremos que a la par con la relación $\inf_{\theta^*} \mathbf{M}_t(\theta^* - t)^2 = 0$, válida para cada t, para la *integral* respecto a $\mathbf{M}_t(\theta^* - t)^2$ ya existe una *frontera inferior positiva que no depende de* θ^* y la cual se halla estrechamente relacionada con la integral análoga de la función $(nI(t))^{-1}$. Así mismo obtendremos, en el caso unidimensional $\theta \in R$, la desigualdad para

$$\inf_{\theta^*} \int \mathbf{M}_t(\theta^* - t)^2 q(t)dt, \tag{1}$$

cualquiera que sea la función ponderal $q(t) \geqslant 0$, $\int q(t)dt = 1$, cuyo segundo miembro no depende de θ^* (incluyendo también el desplazamiento $b(t)$ presente en la desigualdad de Rao—Cramer) y es próximo al valor de J/n, donde

$$J = \int \frac{q(t)}{I(t)}\, dt. \tag{2}$$

2. Desigualdades principales. Antes de enunciar los teoremas respectivos, señalaremos que la integral en (1) puede considerarse como la esperanza matemática incondicional $\mathbf{M}(\theta^* - \theta)^2$ en el caso bayesiano, cuando θ tiene distribución a priori, con una densidad $q(t)$ respecto a la medida de Lebesgue. En este caso $J = \mathbf{M}I^{-1}(\theta)$.

Designemos por $f(x, t) = f_t(x)\, q(t)$ la densidad de la distribución compatible de X, mientras que $\theta \cdot f_t'(x)$, como antes, designará la derivada de $f_t(x)$ respecto a t.

Seguidamente supongamos que $N_h \subset \Theta$ es el portador de la función h definida en Θ: $N_h = \{t: h(t) \neq 0\}$, y que N es el portador de $f(x, t)$ en $\mathscr{X}^n \times \Theta$.

Teorema 1. *Supongamos que $f_t(x)$ es derivable respecto a t, y que la función $\sqrt{I(t)}$ es integrable en cualquier intervalo finito. Entonces para toda función derivable $h(t)$ finita (o sea, igual a 0 fuera del intervalo finito), tal que $N_h \subset N_q$, es válida la desigualdad*

$$\mathbf{M}(\theta^* - \theta)^2 \geqslant \frac{[\mathbf{M}(h(\theta)/q(\theta))]^2}{n\mathbf{M}(I(\theta)[h(\theta)/q(\theta)]^2) + \mathbf{M}[h'(\theta)/q(\theta)]^2} =$$

$$= \frac{\left(\int h(t)dt\right)^2}{n\int I(t)h^2(t)/q(t)dt + \int (h'(t)^2)/q(t)dt}. \qquad (3)$$

Demostración. Tenemos, en virtud del carácter finito de $h(t)$,

$$\int (f_t(x)h(t))'dt = \int d(f_t(x)h(t)) = 0,$$

$$\int t(f_t(x)h(t))'dt = -\int f_t(x)h(t)dt.$$

Por consiguiente, para toda θ^*,

$$\int_{\mathscr{X}^n} \int_{\Theta} (\theta^* - t)(f_t(x)\, h(t))'dt\mu^n(dx) =$$

$$= \int_{\mathscr{X}^n} \int_{\Theta} f_t(x)h(t)dt\mu^n(dx) = \int_{\Theta} h(t)dt. \qquad (4)$$

Estas integrales pueden considerarse, en virtud de la condición $N_h \subset N_q$, como integrales respecto a N. Por lo tanto, podemos multiplicar y dividir por $f(x, t)$ la expresión subintegral en (4). Entonces obtenemos

$$\mathbf{M}\left[(\theta^* - \theta)\, \frac{(f_\theta(X)h(\theta))'}{f(X, \theta)}\right] = \int_{N_q} h(t)dt = \mathbf{M}\, \frac{h(\theta)}{q(\theta)}.$$

De aquí, en virtud de la desigualidad de Cauchy—Buniakovski, resulta

$$\mathbf{M}(\theta^* - \theta)^2 \geqslant \frac{[\mathbf{M}(\mathrm{h}(\theta)/q(\theta))]^2}{\mathbf{M}[(f_\theta(X)h(\theta))'/(f_\theta(X)q(\theta))]^2}. \qquad (5)$$

Sólo queda reducir esta desigualdad a la forma (3). Nótese previamente que

$$\mathbf{M}_t|L'(X,\ t)| \leqslant n\sqrt{I(t)}$$

y que casi para todos*) t,

$$\mathbf{M}_t L'(X,\ t) = 0. \tag{6}$$

La primera de estas afirmaciones se deduce de las relaciones $\mathbf{M}_t|L'(X, t)| \leqslant n\mathbf{M}_t|l'(x_1 t)| \leqslant n\{\mathbf{M}_t[l'(x_1, t)]^2\}^{1/2} = n\sqrt{I(t)}$, que resulta de la desigualdad de Cauchy—Buniakovski. Para demostrar la segunda afirmación tomemos la función finita arbitraria $g(t)$ que en todas partes tiene la derivada continua $g'(t)$. Entonces

$$\int g(t)f_t'(X)dt = -\int g'(t)f_t(X)dt.$$

Además,

$$\int|g(t)|\mathbf{M}_t|L'(X,\ t)|dt \leqslant n\int|g(t)|\sqrt{I(t)}\,dt < \infty.$$

De aquí resulta que se puede cambiar el orden de integración en la expresión siguiente:

$$\int g(t)\mathbf{M}_t L'(X,\ t)dt = \int_{\mathscr{X}}\int_{\Theta} g(t)f_t'(x)dt\mu^n(dx) =$$

$$= -\int_{\mathscr{X}}\int_{\Theta} g'(t)f_t(x)dt\mu^n(dx) = -\int_{\Theta} g'(t)dt = -\int_{\Theta} dg(t) = 0.$$

El cumplimiento de esta igualdad para todos g precisamente significa la validez de (6).

Ahora podemos transformar el segundo miembro (5). Omitiendo, para abreviar, los argumentos de las funciones, obtenemos

$$\mathbf{M}\left[\frac{(f_\theta(X)h(\theta))'}{f_\theta(X)q(\theta)}\right]^2 = \mathbf{M}\left[L'\frac{h}{q} + \frac{h'}{q}\right]^2 = \mathbf{M}\left[\left(\frac{h}{q}\right)^2\mathbf{M}_\theta(L')^2\right] +$$

$$+ 2\mathbf{M}\left[\frac{h'h}{q^2}\mathbf{M}_\theta L'\right] + \mathbf{M}\left(\frac{h'}{q}\right)^2 = n\mathbf{M}\left[\left(\frac{h}{q}\right)^2 I\right] + \mathbf{M}\left(\frac{h'}{q}\right)^2.$$

Aquí hemos aprovechado el hecho de que, en virtud de (6),

$$\mathbf{M}\left[\frac{h'h}{q^2}\mathbf{M}_\theta L'\right] = \int_{N_t}\frac{h'h}{q}\mathbf{M}_t L'dt = 0$$

y que (véase el § 16) $\mathbf{M}_\theta(L')^2 = nI(\theta)$. ◁

En las afirmaciones posteriores siempre supondremos que $f_t(x)$ satisface las condiciones del teorema 1.

---

*) En el § 16 hemos demostrado que esta igualdad, al cumplirse las condiciones (*R*), tiene lugar para todos t. Aquí nos será suficiente que la misma se cumpla para casi todos t.

Teorema 2. *Si la función $h(t) = h_0(t) \equiv q(t)/I(t)$ es finita y derivable, entonces*

$$\mathrm{M}(\theta^* - \theta)^2 \geqslant \left(\frac{J}{n}\left(1 + \frac{H}{nJ}\right)\right)^{-1} \geqslant \frac{J}{n} + \frac{H}{n^2}, \qquad (7)$$

$$donde\ H = \int \left[\left(\frac{q(t)}{I(t)}\right)'\right]^2 \frac{dt}{q(t)}.$$

Observación 1. Las desigualdades dadas en los teoremas 1 y 2 son integrales desde el punto de vista de que pertenecen a las integrales de $\mathrm{M}_t(\theta^* - t)^2$. Desde este punto de vista las desigualdades del § 16 pueden llamarse *locales.*

Demostración. Esta afirmación se deduce directamente del teorema 1, ya que el segundo miembro en (3) se transforma, cuando $h = q/I$, en $J^2/(nJ + H)$. ◁

Por lo tanto, vemos que la frontera inferior de los posibles valores de $\mathrm{M}(\theta^* - \theta)^2$, con grandes valores de n, se distingue poco de la frontera $\frac{J}{n} = \int \frac{q(t)dt}{nI(t)}$ que es igual al valor de $\mathrm{M}(\theta_{\mathfrak{Q}}^* - \theta)^2$ para la estimación R-eficiente $\theta_{\mathfrak{Q}}^*$. Esto muestra que es racional utilizar las estimaciones eficientes, puesto que para ellas, cualquiera que sea la función q, casi se alcanza el valor extremal de $\mathrm{M}(\theta^* - \theta)^2$.

La estimación (7) es inmejorable, lo cual es confirmado por el

Ejemplo 1. Sea $X \in \Phi_{\alpha,1}$. Como sabemos, en este caso $I(\alpha) = 1$. Supongamos, luego, que el parámetro α se elige aleatoriamente con una densidad suave de $q(t)$, $t \in (-\infty, \infty)$. Entonces el segundo miembro de (7) se transforma en $(n + H)^{-1}$, donde

$$H = \int \frac{(q')^2}{q} dt = \mathrm{M}[(\ln q(\alpha))']^2.$$

Es nuestro caso, la estimación bayesiana $\alpha_{\mathfrak{Q}}^*$, que corresponde a la distribución a priori $\mathbf{Q}$ con densidad q y que minimiza $\mathrm{M}(\alpha^* - \alpha)^2$, es igual a (véase el § 10)

$$\alpha_{\mathfrak{Q}}^* = \frac{\int tq(t)f_t(X)dt}{\int q(t)f_t(X)dt} =$$

$$= \frac{\int tq(t)\exp(n\bar{x}t - t^2n/2)dt}{\int q(t)\exp(n\bar{x}t - t^2n/2)dt} = \frac{\int tq(t)\exp(-n(\bar{x} - t)^2/2)dt}{\int q(t)\exp(-n(\bar{x} - t)^2/2)dt}. \qquad (8)$$

Es fácil hallar la representación asintótica de esta relación y mostrar que

$$\alpha_{\mathfrak{Q}}^* = \bar{x} + \frac{q'(\bar{x})}{nq(\bar{x})} + O\left(\frac{1}{n^2}\right), \quad \mathrm{M}(\alpha_{\mathfrak{Q}}^* - \alpha)^2 = \frac{1}{n} - \frac{H}{n^2} + O\left(\frac{1}{n^3}\right).$$

No obstante, procederemos más sencillamente, suponiendo que $q(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$.

Entonces es evidente que $H = 1$, y el segundo miembro en (7) se convierte en $1/(n + 1)$. Pero en el ejemplo 11.1 hemos establecido que

$$\mathrm{M}(\alpha_Q^* - \alpha)^2 = \frac{1}{n+1}.$$

De este modo, la inmejorabilidad de las igualdades (7) y (3) queda demostrada.

Teorema 3. *Si el intervalo* $(a - \varepsilon, a + \varepsilon)$ *se contiene en* Θ, *entonces, para toda estimación* θ^*,

$$\max_{t \in (a-\varepsilon, a+\varepsilon)} \mathrm{M}_t(\theta^* - t)^2 \geqslant \frac{1}{n \max\limits_{t \in (a-\varepsilon, a+\varepsilon)} I(t) + \pi^2 \varepsilon^{-2}}.$$

Demostración. Hagamos uso de la desigualdad

$$\max_{t \in (a-\varepsilon, a+\varepsilon)} \mathrm{M}_t(\theta^* - t)^2 \geqslant \int_{a-\varepsilon}^{a+\varepsilon} \mathrm{M}_t(\theta^* - t)^2 q(t)\,dt,$$

válida para toda densidad $q(t)$ que es igual a cero fuera de $(a - \varepsilon, a + \varepsilon)$. La afirmación necesaria se deduce del teorema 1 si suponemos en éste

$$h(t) = q(t) = \frac{1}{\varepsilon} \cos^2 \frac{\pi(t-a)}{2\varepsilon}, \quad |t - a| \leqslant \varepsilon.$$

Entonces

$$\mathrm{M}_t(\theta^* - \theta)^2 \geqslant \frac{1}{n\int I(t)q(t)\,dt + \int (q'(t))^2/q(t)\,dt},$$

donde

$$\int \frac{(q'(t))^2}{q(t)}\,dt = \int_{-\varepsilon}^{\varepsilon} \frac{\left(\frac{\pi}{2\varepsilon^2}\, 2\cos\frac{\pi t}{2\varepsilon} \operatorname{sen}\frac{\pi t}{2\varepsilon}\right)^2 \varepsilon}{\cos^2 \frac{\pi t}{2\varepsilon}}\,dt =$$

$$= \frac{1}{\varepsilon^2} \int_{-1}^{1} \pi^2 \operatorname{sen}^2 \frac{\pi t}{2}\,dt = \frac{\pi^2}{\varepsilon^2}. \quad \triangleleft$$

Se puede señalar que en la función $q(t) = \cos^2(\pi t/2)$ se alcanza el mínimo de la funcional $\int_{-1}^{1} (q'(t))^2/q(t)\,dt$ en la clase de todas las densidades derivables $q(t)$.

Del teorema 3 se deduce, en particular, que el intervalo de valores de θ para los cuales la estimación θ^* es supereficiente no puede tener una longitud mayor que $O(1/\sqrt{n})$.

3. Desigualdades en el caso cuando la función $q(\theta)/I(\theta)$ no es derivable. Si la función $h_0 = q/I$ no satisface las condiciones del teorema 1, es válida la siguiente afirmación útil que permite estimar la asintótica de $\mathbf{M}(\theta^* - \theta)^2$ en el caso general.

Teorema 4. *Supongamos que la sucesión de funciones $h_\varepsilon(t)$, dependientes del parámetro $\varepsilon > 0$, es tal que cada función h satisface las condiciones del teorema 1 y*

1) $h_\varepsilon(t) \leqslant h_0(t)$,

2) $H(\varepsilon) = \int \frac{(h'_\varepsilon(t))^2}{q(t)}\, dt < \infty$.

Entonces, para todo $\varepsilon > 0$,

$$\mathbf{M}(\theta^* - \theta)^2 \geqslant \frac{\left(\int h_\varepsilon(t)dt\right)^2}{nJ + H(\varepsilon)}.$$

La demostración se deduce directamente del teorema 1 si se toma $h = h_\varepsilon$.

Del teorema 4 obtenemos el siguiente colorario importante.

Teorema 5. *Si la función q es integrable según Riemann, $J < \infty$, entonces*

$$\mathbf{M}(\theta^* - \theta)^2 \geqslant \frac{J}{n}(1 + \delta_n),$$

donde $\delta_n = o(1)$ cuando $n \to \infty$.

Demostración. Pongamos $\hat{q}_\varepsilon(t) = \min\limits_{|u| \leqslant \varepsilon} q(t + u)$,

$$q_\varepsilon(t) = \begin{cases} \hat{q}_\varepsilon(t) & si\ \hat{q}_\varepsilon(t) \geqslant \varepsilon, \\ 0 & si\ \hat{q}_\varepsilon(t) < \varepsilon, \end{cases}$$

$$I_\varepsilon(t) = \text{máx}(\varepsilon, I(t)),$$

$$h_\varepsilon(t) = \frac{1}{2\varepsilon} \int_{t-\varepsilon}^{t+\varepsilon} \frac{q_\varepsilon(v)}{I_\varepsilon(v)}\, dv \leqslant h_0(t).$$

Es evidente que la función h_ε es finita y derivable para cualquier $\varepsilon > 0$.

Del hecho de que $q(t)$ es integrable según Riemann se desprende que $q_\varepsilon(t) \uparrow q(t)$ casi en todas las partes cuando $\varepsilon \to 0$. Para demostrar esto cerciorémonos de que

$$\int_a^b [q(t) - q_\varepsilon(t)]dt \downarrow 0. \qquad (9)$$

De la integrabilidad de $q(t)$ según Riemann se deduce la convergencia

$$\sum_k q_\delta(2k\delta)2\delta \uparrow \int q(t)dt,$$

$$\sum_k q_\delta((2k + 1)\delta)2\delta \uparrow \int q(t)dt$$

cuando $\delta \to 0$. Por eso

$$\int_a^b q_\varepsilon(t)dt \geqslant \sum_k q_{2\varepsilon}(2k\varepsilon)2\varepsilon =$$

$$= \frac{1}{2}\left(\sum q_{2\varepsilon}(4k\varepsilon)4\varepsilon + \sum q_{2\varepsilon}((4k + 2)\varepsilon)4\varepsilon\right) \to \int q(t)dt.$$

La relación (9), y junto con ella la convergencia de $q_\varepsilon(t) \uparrow q(t)$, quedan demostradas.

Utilizando ahora esta convergencia, obtenemos $\frac{q_\varepsilon(t)}{I_\varepsilon(t)} \uparrow h_0(t)$,

$$\int h_\varepsilon(t)dt = \int \frac{dt}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} \frac{q_\varepsilon(t + v)}{I_\varepsilon(t + v)}dv =$$

$$= \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} dv \int \frac{q_\varepsilon(t)}{I_\varepsilon(t)}\, dt = \int \frac{q_\varepsilon(t)}{I_\varepsilon(t)}\, dt \uparrow J.$$

Además,

$$|h_\varepsilon'(t)| = \frac{1}{2\varepsilon}\left|\frac{q_\varepsilon(t + \varepsilon)}{I_\varepsilon(t + \varepsilon)} - \frac{q_\varepsilon(t - \varepsilon)}{I_\varepsilon(t - \varepsilon)}\right| \leqslant \frac{q(t)}{\varepsilon^2},$$

$$H(\varepsilon) \leqslant \int \left(\frac{q(t)}{\varepsilon^2}\right)^2 q^{-1}(t)dt = \frac{1}{\varepsilon^4}.$$

Ahora podemos hacer uso del teorema 3. Suponiendo $\varepsilon = \varepsilon(n) = n^{-1/5}$, $n \to \infty$, obtenemos $\varepsilon(n) \to 0$,

$$\mathbf{M}(\theta^* - \theta)^2 \geqslant \frac{\left(\int h_\varepsilon(t)dt\right)^2}{nJ + n^{4/5}} = \frac{J}{n}(1 + o(1)). \quad \triangleleft$$

4. Algunos corolarios. Criterios del carácter asintóticamente bayesiano y minimax. Una de las principales conclusiones que pueden sacarse de los resultados de este párrafo consiste, hablando en general, en lo siguiente. Si existe la estimación asintóticamente R-eficiente, cualquiera que sea otra estimación que tomemos, no obtendremos "en total" (o "por término medio") un resultado asintóticamente mejor. Utilicemos este hecho, más tarde, en el § 25. Aquí sólo expondremos los criterios del carácter asintóticamente bayesiano y del carácter asintóticamente minimax que se desprenden directamente de los teoremas 2 y 5.

Definición 1. La estimación θ_1^*, que posee la propiedad

$$\mathbf{M}n(\theta_1^* - \theta)^2 = J + o(1) \tag{10}$$

cuando $n \to \infty$, se llama *R-bayesiana asintóticamente*.

Son las estimaciones para las cuales se alcanza asintóticamente la frontera inferior de las desviaciones estándar, definida en los teoremas 2.5. Las

mismas también podrían denominarse *estimaciones R-eficientes "en total"* (o "por término medio").

Recordemos (véase el § 11) que la estimación θ_1^* se llama *asintóticamente bayesiana* (con respecto a la distribución **Q**) si para cualquier otra estimación θ^*

$$\lim_{n\to\infty} \sup\, [\mathbf{M}n(\theta_1^* - \theta)^2 - \mathbf{M}n(\theta^* - \theta)^2] \leqslant 0. \tag{11}$$

Corolario 1. *Supongamos que se cumplen las condiciones del teorema 1 y que la función $q(t)$ es integrable según Riemann. Entonces una estimación asintóticamente R-bayesiana es asintóticamente bayesiana.*

Demostración. Supongamos que θ_1^* es una estimación asintóticamente R-bayesiana. En virtud del teorema 5, para toda estimación θ^*,

$$\lim_{n\to\infty} \inf \mathbf{M}n(\theta^* - \theta)^2 \geqslant J.$$

De aquí y de (10) resulta (11).

También está claro que si existe una estimación asintóticamente R-bayesiana, toda estimación asintóticamente bayesiana será R-bayesiana (compárese con las observaciones referentes al teorema 16.3).

Del teorema 5 también se desprende el

Corolario 2. *Supongamos que se cumplen las condiciones del teorema 1 y que la función $q(t)$ es integrable según Riemann. Si θ_1^* y θ_2^* son dos estimaciones asintóticamente R-bayesianas, éstas son asintóticamente equivalentes desde el punto de vista siguiente:*

$$\mathbf{M}n(\theta_1^* - \theta_2^*)^2 \to 0, \quad (\theta_1^* - \theta_2^*)\sqrt{n} \xrightarrow[P]{} 0,$$

donde la convergencia en probabilidad se entiende respecto a la distribución compatible de X y θ en $\mathscr{X}^n \times \Theta$.

La demostración es completamente análoga a las demostraciones de los teoremas 8.2, 16.4. La igualdad inicial (8.11), en virtud del teorema 5, da

$$\lim_{n\to\infty} \sup \mathbf{M}n(\theta_1^* - \theta_2^*)^2 \leqslant 0. \quad \triangleleft$$

En los §§ 8 y 11 hemos señalado que para comparar las estimaciones, a la par con los valores *medios* $\int q(t)\mathbf{M}_t(\theta^* - t)^2 dt$, pueden considerarse los valores máximos

$$\sup_{t\in\Gamma} \mathbf{M}_t(\theta^* - t)^2, \quad \Gamma \subset \Theta.$$

En calidad de Γ se toma todo el conjunto Θ o la parte de éste que, según datos previos, contiene el valor desconocido de θ. Recordemos que la estimación $\bar{\theta}^*$ se llama minimax cuando para toda estimación θ^*

$$\sup_{t\in\Gamma} \mathbf{M}_t(\bar{\theta}^* - t)^2 \leqslant \sup_{t\in\Gamma} \mathbf{M}_t(\theta^* - t)^2.$$

La estimación θ_1^* se llama *asintóticamente minimax* cuando para toda estimación θ^*

$$\lim_{n\to\infty}\sup \sup_{t\in\Gamma} \mathbf{M}_t[\sqrt{n}(\theta_1^* - t)]^2 \leqslant \lim_{n\to\infty}\inf \sup_{t\in\Gamma} \mathbf{M}_t[\sqrt{n}(\theta^* - t)]^2.$$

Corolario 3. *Supongamos que la información de Fisher $I(\theta)$ existe y es continua. En este caso, si para cualquier segmento $\Gamma \subset \Theta$,*

$$\lim_{n\to\infty}\sup \sup_{t\in\Gamma} \mathbf{M}_t[\sqrt{n}(\theta_1^* - t)]^2 \leqslant \sup_{t\in\Gamma} I^{-1}(t), \tag{12}$$

entonces la estimación θ_1^* es asintóticamente minimax.

Demostración. Es suficiente convencerse de que para cualquier estimación θ^*,

$$\lim_{n\to\infty}\inf \sup_{t\in\Gamma} \mathbf{M}_t[\sqrt{n}(\theta^* - t)]^2 \geqslant \sup_{t\in\Gamma} I^{-1}(t). \tag{13}$$

Para cualquier distribución $\mathbf{Q}$ en Γ, con una densidad suave $q(t)$ respecto a la medida de Lebesgue,

$$\sup_{t\in\Gamma} \mathbf{M}_t[\sqrt{n}(\theta^* - t)]^2 \geqslant \int \mathbf{M}_t[\sqrt{n}(\theta^* - t)]^2 q(t)dt.$$

Según el teorema 2, la integral del segundo miembro es para cualquier estimación θ^*, no menor que $J - H/n$. Por eso el primer miembro de (13) es mayor o igual a

$$J = \int I^{-1}(t)q(t)dt.$$

Pero q es una densidad suave arbitraria y, para un valor dado de $\varepsilon > 0$, la misma siempre se puede elegir, en virtud de la continuidad de $I^{-1}(t)$, de modo que

$$J \geqslant \sup_{t\in\Gamma} I^{-1}(t) - \varepsilon.$$

En vista de que ε es arbitrario, (13) queda demostrada. ◁

En conclusión de este apartado es necesario hacer una observación importante, que consiste en que, al buscar las estimaciones asintóticamente óptimas, es posible limitarse a la clase $\mathfrak{K}_0$ de estimaciones asintóticamente no desplazadas, que hemos introducido en el § 16. Esto se deduce de las consideraciones siguientes.

Ya hemos señalado que el segundo miembro de la desigualdad del teorema 5, equivalente a $J/n + o(1/n)$, no depende absolutamente del desplazamiento $b(\theta)$. Al mismo tiempo, si al construir la frontera inferior de $\mathbf{M}(\theta^* - \theta)^2$ partimos de la desigualdad de Rao—Cramer dada en el § 16, entonces obtendremos

$$\mathbf{M}(\theta^* - \theta)^2 \geqslant \min_b \int q(t)\left[\frac{(1 + b'(t))^2}{nI(t)} + b^2(t)\right]dt.$$

Se puede mostrar (compárese con [47]) que este valor mínimo de todos los desplazamientos $b(\theta)$ tiene (con ciertas suposiciones acerca de la suavidad de $q(t)$ y $I(t)$) esa misma forma $J/n + o(1/n)$ y (lo cual es esencial para nosotros) se alcanza en el desplazamiento $b(\theta)$ que posee, cuando $n \to \infty$, las propiedades

$$b'(t) = o(1), \; b(t) = o(1/\sqrt{n}).$$

La clase de estimaciones θ^* con tales desplazamientos es precisamente $\tilde{K}_0$ (véase el § 16). La salida de θ^* de la clase $\tilde{K}_0$ hace inaccesible la frontera $J/n + o(1/n)$. Ahora bien, en el enfoque asintótico, cuando las estimaciones asintóticamente normales se comparan con ayuda de los valores de $\mathbf{M}(\theta^* - \theta)^2$ cuando son suaves $q(t)$ e $I(t)$, es posible limitarse a examinar las estimaciones de la clase $K = K_{\Phi,2} \cap \tilde{K}_0$ (hemos examinado la clase $K_{\Phi,2}$ en el § 8), puesto que las estimaciones fuera de la clase $\tilde{K}_0$ son "inadmisibles" desde el punto de vista antes indicado.

5. **Caso multidimensional.** En el caso de $\theta \in R^k$ se pueden obtener los análogos para todos los teoremas de este párrafo y hacer las mismas deducciones que hemos obtenido para el caso unidimensional.

En particular, la afirmación del teorema 5, uno de los principales en este apartado, tendrá la forma

$$d^2 \geqslant J/n + o(1/n),$$

donde $d^2 = \|d_{ij}\|$, $d_{ij} = \mathbf{M}(\theta_i^* - \theta_i)(\theta_j^* - \theta_j)$, $J = \mathbf{M}I^{-1}(\theta)$.

Los razonamientos relacionados con las estimaciones bayesianas y minimax también conservan su validez cuando en calidad de error de la estimación se considera el valor

$$v(\theta^*) = \mathbf{M}_\theta(\theta^* - \theta)V(\theta^* - \theta)^T,$$

donde V es una matriz definida no negativamente. Deben llamarse estimaciones bayesianas o minimax (o asintóticamente bayesianas y minimax) las estimaciones cuyos errores satisfacen las desigualdades respectivas para cualquier matriz V definida no negativamente.

§ 21. Distancias de Kullback—Leibler, de Hellinger y χ^2. Sus propiedades

Los resultados de este párrafo serán esenciales para la obtención de los resultados principales de la teoría asintótica de estimación, así como para los resultados del cap. 3.

1. **Definiciones y propiedades principales de las distancias.**

Sean $\mathbf{P}$ y $\mathbf{G}$ dos distribuciones en $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}})$ absolutamente continuas respecto a la medida μ. Designemos

$$\frac{d\mathbf{P}}{d\mu} = p, \; \frac{dG}{d\mu} = g,$$

N_p es el portador de la distribución $\mathbf{P}$: $N_p = \{x: p(x) > 0\}$.

Definición 1. Se llama *distancia de Kullback—Leibler* entre las distribuciones **P** y **G** la magnitud

$$\varrho_1(\mathbf{P}, \mathbf{G}) = \int_{N_p} \ln \frac{p(x)}{g(x)} \mathbf{P}(dx) = \int_{N_p} \ln \frac{p(x)}{g(x)} p(x)\mu(dx).$$

De hecho $\varrho_1(\mathbf{P}, \mathbf{G})$ no es, por supuesto, una distancia o una métrica en sentido general, ya que $\varrho_1(\mathbf{P}, \mathbf{G})$ no es una función simétrica de **P** y **G**. No obstante, veremos que $\varrho_1(\mathbf{P}, \mathbf{G})$ caracteriza en realidad (desde el punto de vista estadístico) la desviación de **G** respecto a **P**.

De la desigualdad $\ln(1 + v) - v \leqslant 0$ y la representación

$$\varrho_1(\mathbf{P}, \mathbf{G}) = -\int \left[\ln \frac{g}{p} - \left(\frac{g}{p} - 1\right)\right] p\mu(dx)$$

se deduce que siempre $\varrho_1(\mathbf{P}, \mathbf{G}) \geqslant 0$. En el lema 6.1 hemos establecido que la desigualdad $\varrho_1(\mathbf{P}, \mathbf{G}) = 0$ sólo es posible si $\mathbf{P} = \mathbf{G}$.

Definición 2. Llamaremos *distancia* χ^2 entre las distribuciones **P** y **G** la magnitud

$$\varrho_2(\mathbf{P}, \mathbf{G}) = \int_{N_p \cup N_0} \frac{(p(x) - g(x))^2}{p(x)} \mu(dx).$$

Casi todas las observaciones hechas para la definición 1 se refieren a esta distancia. La denominación de χ^2 se explica por razones que serán aclaradas más tarde.

Definición 3. Se llama *distancia de Hellinger* entre las distribuciones **P** y **G** la magnitud

$$\varrho_3(\mathbf{P}, \mathbf{G}) = \int_{N_p \cup N_0} \left(\sqrt{p(x)} - \sqrt{g(x)}\right)^2 \mu(dx).$$

La distancia de Hellinger ya es la función simétrica de **P** y **G**, y el valor de $\sqrt{\varrho_3(\mathbf{P}, \mathbf{G})}$ posee todas las propiedades de la métrica (entre las funciones $\sqrt{p(x)}$ y $\sqrt{g(x)}$ en el espacio métrico $L_2(\mathscr{X}, \mu)$). Es fácil notar que

$$\varrho_3(\mathbf{P}, \mathbf{G}) = 2\left(1 - \int \sqrt{pq}\, \mu(dx)\right) \leqslant 2. \quad (1)$$

Las tres distancias introducidas desempeñan un papel importante en distintos problemas de la estadística matemática. Nos convenceremos de ello en cierta medida.

Si mediante estas distancias se caracteriza el grado de proximidad de las distribuciones, cuando la relación p/g es próxima a 1, resultará que todas ellas se comportan asintóticamente igual, con una exactitud de hasta los factores constantes. En efecto, valiéndose del desarrollo

$$\ln \frac{g}{p} = \ln\left(1 + \left(\frac{g}{p} - 1\right)\right) = \left(\frac{g}{p} - 1\right) - \frac{1}{2}\left(\frac{g}{p} - 1\right)^2 + O\left(\left|\frac{g}{p} - 1\right|\right)^3,$$

obtenemos

$$\varrho_1(\mathbf{P},\ \mathbf{G}) = -\int \ln\frac{g}{p}\cdot p\mu(dx) \approx \frac{1}{2}\int\left(\frac{g}{p}-1\right)^2 p\mu(dx) = \frac{1}{2}\,\varrho_2(\mathbf{P},\ \mathbf{G}),$$

$$\varrho_2(\mathbf{P},\ \mathbf{G}) = \int\frac{(p-g)^2}{p}\,\mu(dx) = \int(\sqrt{p}-\sqrt{g})^2\left(1+\sqrt{\frac{g}{p}}\right)^2\mu(dx) \approx$$

$$\approx 4\varrho_3(\mathbf{P},\ \mathbf{G}).$$

De la última igualdad también se deduce que $\varrho_2(\mathbf{P},\ \mathbf{G}) \geqslant \varrho_3(\mathbf{P},\ \mathbf{G})$.

Además, $\varrho_1(\mathbf{P},\ \mathbf{G}) \geqslant \varrho_3(\mathbf{P},\ \mathbf{G})$. En efecto, como $\ln(1+x) \leqslant x$, entonces

$$\ln\frac{g}{p} = 2\ln + \left(1+\left(\sqrt{\frac{g}{p}}-1\right)\right) \leqslant 2\left(\sqrt{\frac{g}{p}}-1\right),$$

$$\varrho_1(\mathbf{P},\ \mathbf{G}) = -\int\ln\frac{g}{p}\,p\mu(dx) \geqslant -2\left(\int\sqrt{pg}\,\mu(dx)+1\right) = \varrho_3(\mathbf{P},\ \mathbf{G}).$$

En lo sucesivo examinaremos el caso paramétrico y consideraremos que se cumple la condición $(A\mu)$. Nos interesarán las distancias ϱ_i, $i = 1, 2, 3$, entre las distribuciones $\mathbf{P} = \mathbf{P}_{\theta_1}$ y $\mathbf{G} = \mathbf{P}_{\theta_2}$ en $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}})$, así como entre las distribuciones muestrales correspondientes (aquí las designaremos por $\mathbf{P}^n_{\theta_1}$, $\mathbf{P}^n_{\theta_2}$) en $(\mathscr{X}^n, \mathfrak{B}^n_{\mathscr{X}})$. (Señalemos que las distancias tienen sentido para las distribuciones arbitrarias, y con la naturaleza de los espacios no están relacionadas de ningún modo). Si $N_{P_{\theta_2}} \subset N_{P_{\theta_1}}$, podemos escribir

$$\varrho_1(\mathbf{P}_{\theta_1},\ \mathbf{P}_{\theta_2}) = \int\ln\frac{f_{\theta_1}}{f_{\theta_2}}\,f_{\theta_1}\mu(dx) = \mathbf{M}_{\theta_1}\ln\frac{f_{\theta_1}(\mathrm{x}_1)}{f_{\theta_2}(\mathrm{x}_1)},$$

$$\varrho_2(\mathbf{P}_{\theta_1},\ \mathbf{P}_{\theta_2}) = \int\frac{(f_{\theta_1}-f_{\theta_2})^2}{f_{\theta_1}}\,\mu(dx) = \mathbf{M}_{\theta_1}\left(\frac{f_{\theta_2}(\mathrm{x}_1)}{f_{\theta_1}(\mathrm{x}_1)}-1\right)^2, \qquad (2)$$

$$\varrho_3(\mathbf{P}_{\theta_1},\ \mathbf{P}_{\theta_2}) = \int\left(\sqrt{f_{\theta_1}}-\sqrt{f_{\theta_2}}\right)^2\mu(dx) = \mathbf{M}_{\theta_1}\left(\sqrt{\frac{f_{\theta_2}(\mathrm{x}_1)}{f_{\theta_1}(\mathrm{x}_1)}}-1\right)^2.$$

Si no se cumple la condición $N_{P_{\theta_2}} \subset N_{P_{\theta_1}}$ entonces $\varrho_2(\mathbf{P}_{\theta_1},\ \mathbf{P}_{\theta_2})$, $\varrho_3(\mathbf{P}_{\theta_1},\ \mathbf{P}_{\theta_2})$ serán mayores que las esperanzas matemáticas correspondientes en (2).

Cabe señalar que a la par con (2) tiene lugar la siguiente igualdad útil que se desprende de (1):

$$\mathbf{M}_{\theta_1}\sqrt{f_{\theta_2}(\mathrm{x}_1)/f_{\theta_1}(\mathrm{x}_1)} = \int\sqrt{f_{\theta_2}(x)f_{\theta_1}(x)}\,\mu(dx) =$$

$$= 1 - \frac{1}{2}\,\varrho_3(\mathbf{P}_{\theta_1},\ \mathbf{P}_{\theta_2}). \qquad (3)$$

La relación entre las distancias $\varrho_i(\mathbf{P}_{\theta_1},\ \mathbf{P}_{\theta_2})$ y $\varrho_i(\mathbf{P}^n_{\theta_1},\ \mathbf{P}^n_{\theta_2})$ se establece por la afirmación siguiente.

Teorema 1.

$$\begin{gathered}\varrho_1(\mathbf{P}^n_{\theta_1}, \mathbf{P}^n_{\theta_2}) = n\varrho_1(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}),\\ 1 + \varrho_2(\mathbf{P}^n_{\theta_1}, \mathbf{P}^n_{\theta_2}) = (1 + \varrho_2(\mathbf{P}_{\theta_1}, \mathbf{P}^n_{\theta_2}))^n,\\ 1 - \frac{1}{2}\varrho_3(\mathbf{P}^n_{\theta_1}, \mathbf{P}^n_{\theta_2}) = \left(1 - \frac{1}{2}\varrho_3(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2})\right)^n.\end{gathered} \tag{4}$$

La demostración es casi evidente si se supone, para abreviar, que $N_{P_{\theta_2}} \subset N_{P_{\theta_1}}$ (en el caso general los cálculos conservarán, de hecho, su validez, pero serán un poco más voluminosos). En efecto, en este caso podemos hacer uso de las igualdades (2). Entonces la primera de las relaciones (4) se deduce directamente del hecho de que

$$\ln\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} = \sum_{i=1}^{n} \ln\frac{f_{\theta_1}(x_i)}{f_{\theta_2}(x_i)}.$$

Seguidamente, en virtud de (2),

$$\begin{gathered}1 + \varrho_2(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}) = \mathbf{M}_{\theta_1}(f_{\theta_2}(x_1)/f_{\theta_1}(x_1))^2,\\ 1 - \varrho_3(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2})/2 = \mathbf{M}_{\theta_1}\sqrt{f_{\theta_2}(x_1)/f_{\theta_1}(x_1)},\end{gathered}$$

y las relaciones de este mismo tipo son válidas para las distancias entre $\mathbf{P}^n_{\theta_1}$ y $\mathbf{P}^n_{\theta_2}$ (sustituyendo en los segundos miembros x_1 por X). Como

$$\mathbf{M}_{\theta_1}\left(\frac{f_{\theta_2}(X)}{f_{\theta_1}(X)}\right)^\alpha = \mathbf{M}_{\theta_1}\prod_{i=1}^{n}\left(\frac{f_{\theta_2}(x_i)}{f_{\theta_1}(x_i)}\right)^\alpha = \left[\mathbf{M}_{\theta_1}\left(\frac{f_{\theta_2}(x_1)}{f_{\theta_1}(x_1)}\right)^\alpha\right]^n,$$

de aquí, cuando $\alpha = 2$ y $\alpha = 1/2$, obtenemos (4).

Le recomendamos al lector que demuestre este teorema en el caso general (o sea, cuando no se cumple la condición $N_{P_{\theta_2}} \subset N_{P_{\theta_1}}$). ◁

Del teorema 1 se desprende el

Corolario 1.

$$\varrho_3(\mathbf{P}^n_{\theta_1}, \mathbf{P}^n_{\theta_2}) \leqslant n\varrho_3(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}).$$

En efecto, $1 - \beta^n \leqslant (1 - \beta)n$ para cualquier $\beta \geqslant 0$. Suponiendo $\beta = 1 - \frac{1}{2}\varrho_3(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2})$, obtenemos de (4),

$$\varrho_3(\mathbf{P}^n_{\theta_1}, \mathbf{P}^n_{\theta_2}) = 2(1 - \beta^n) \leqslant 2(1 - \beta)n = n\varrho_3(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}). \quad ◁$$

2. Relación de las distancias de Hellinger y otras con la información de Fisher. Entre las tres distancias introducidas en el apartado anterior, en lo sucesivo, la distancia de Hellinger tendrá para nosotros, el mayor interés. Al mismo tiempo, el carácter de las afirmaciones principales, expuestas más abajo (teoremas 2 y 3), y el carácter de las demostraciones serán iguales para las tres distancias. Por eso, para abreviar, nos limitare-

mos, en este apartado, a estudiar la distancia de Hellinger, que designaremos (omitiendo el índice del símbolo ϱ_3) del siguiente modo:

$$\varrho(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}) = \int \left(\sqrt{f_{\theta_1}} - \sqrt{f_{\theta_2}}\right)^2 \mu(dx).$$

Pongamos $r(\theta_1, \theta_2) = \varrho(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2})$.

Lema 1. *Si $f_\theta(x)$, para c.t. $[\mu]$ valores de x, es continua respecto a θ, $\theta_1 \neq \theta_2$, entonces*

$$\lim_{\substack{\theta' \to \theta_1 \\ \theta'' \to \theta_2}} \inf \frac{r(\theta', \theta'')}{|\theta' - \theta''|^2} \geqslant \frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2}. \tag{5}$$

Si la función $\sqrt{f_\theta(x)}$, para c.t. $[\mu]$ valores de x, es derivable respecto a θ, entonces

$$\lim_{\substack{\theta' \to \theta \\ \theta'' \to \theta}} \inf \frac{r(\theta', \theta'')}{|\theta' - \theta''|^2} \geqslant \frac{I(\theta)}{4}. \tag{6}$$

Además,

$$\frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \leqslant \frac{1}{4} \int_0^1 I(\theta_1 + (\theta_2 - \theta_1)y)dy. \tag{7}$$

Aquí se supone, claro está, que los valores de θ', θ'', θ_1, θ_2, θ pertenecen a Θ.

Demostración. Para verificar (5) es suficiente utilizar el lema de Fatou y la continuidad de $f_\theta(x)$ en la relación

$$\lim_{\substack{\theta' \to \theta_1 \\ \theta'' \to \theta_2}} \inf \frac{r(\theta', \theta'')}{|\theta' - \theta''|^2} \geqslant \int \lim_{\substack{\theta' \to \theta_1 \\ \theta'' \to \theta_2}} \inf \left(\frac{\sqrt{f_{\theta'}} - \sqrt{f_{\theta''}}}{\theta' - \theta''}\right)^2 \mu(dx).$$

En vista de que, cuando $\theta_1 = \theta_2 = \theta$, la expresión subintegral en la última integral es igual a $(f'_\theta)^2/(4f_\theta)$, obtenemos (6).

Para demostrar (7) pongamos $a = \theta_2 - \theta_1$ y representemos el incremento $\sqrt{f_{\theta_2}} - \sqrt{f_{\theta_1}}$ en la forma

$$\frac{1}{2} \int_{\theta_1}^{\theta_2} \frac{f'_t}{\sqrt{f_t}} dt = \frac{a}{2} \int_0^1 \frac{f'_{\theta_1 + ay}}{\sqrt{f_{\theta_1 + ay}}} dy.$$

En virtud de la desigualdad de Cauchy—Buniakovski,

$$(\sqrt{f_{\theta_2}} - \sqrt{f_{\theta_1}})^2 = \left[\frac{a^2}{4} \int_0^1 \frac{f'_{\theta_1 + ay}}{\sqrt{f_{\theta_1 + ay}}} dy\right]^2 \leqslant \frac{a^2}{4} \int_0^1 \frac{(f'_{\theta_1 + ay})^2}{f_{\theta_1 + ay}} dy.$$

Utilizando la negatividad de la función subintegral, podemos cambiar el

orden de integración en las relaciones siguientes:

$$\frac{r(\theta_1, \theta_2)}{a^2} \leqslant \frac{1}{4} \int_{\mathscr{X}} \left(\int_0^1 \frac{(f'_{\theta_1+ay})^2}{f_{\theta_1+ay}} dy \right) \mu(dx) = \frac{1}{4} \int_0^1 I(\theta_1 + ay) dy.$$

La desigualdad (7) queda demostrada. ◁

Pongamos $r(\Delta) = r(\theta, \theta + \Delta)$. Del lema 1 se deduce directamente el

Teorema 2. *Si la función $\sqrt{f_\theta(x)}$, para c.t. $[\mu]$ valores de x, es derivable respecto a θ, e $I(\theta)$ es continua, entonces existe*

$$\lim_{\Delta \to 0} \frac{r(\Delta)}{\Delta^2} = \frac{I(\theta)}{4}. \tag{8}$$

Observación 1. Esta afirmación también será válida para las distancias ϱ_1 y ϱ_2 si suponemos

$$r(\Delta) = \frac{1}{4} \varrho_2(\mathbf{P}_\theta, \mathbf{P}_{\theta+\Delta}), \ r(\Delta) = \frac{1}{2} \varrho_1(\mathbf{P}_\theta, \mathbf{P}_{\theta+\Delta}).$$

En este caso, la relación (6) se demuestra exactamente igual que en el lema 1. La demostración de (8) puede exigir la utilización de condiciones adicionales de regularidad (próximas a las condiciones (R)) que aseguren la validez del paso límite bajo el signo integral.

Así pues, $\varrho_i(\mathbf{P}_\theta, \mathbf{P}_{\theta+\Delta})$, $i = 1, 2, 3$, se comportan asintóticamente igual, e $I(\theta)$ caractariza la velocidad de su tendencia hacia el cero cuando $\Delta \to 0$ $\left(\text{pues } \frac{1}{4} I(\theta) \text{ es la segunda derivada de } r(v) \text{ en el punto } v = 0\right)$.

Si se pone $r^{(n)}(\Delta) = \varrho(\mathbf{P}^n_{\theta+\Delta}, \mathbf{P}^n_\theta)$, entonces, de los teoremas 1 y 2 resultará

$$\lim_{\Delta \to 0} \frac{r^{(n)}(\Delta)}{\Delta^2} = \frac{nI(\theta)}{4}.$$

Estas mismas relaciones se mantendrán para las distancias ϱ_1 y ϱ_2.

3. Existencia de fronteras uniformes para $r(\Delta)/\Delta^2$. En lo sucesivo, la existencia de tales fronteras nos permitirá obtener estimaciones muy útiles para los momentos de relación de verosimilitud.

A fin de simplificar la exposición o evitar la introducción de otras condiciones más voluminosas, en las investigaciones posteriores a menudo estimaremos que se cumple la condición

(A_c): *el conjunto Θ es compacto.*

Desde el punto de vista de las aplicaciones, esta condición, que significa el carácter limitado y cerrado del conjunto paramétrico, por lo general, no es limitativa.

Más adelante también utilizaremos la condición (A_0) que hemos introducido en el § 6 y que significa que $f_{\theta_1} \neq f_{\theta_2}$ cuando $\theta_1 \neq \theta_2$. Con esta condición, $r(\theta_1, \theta_2) > 0$ cuando $\theta_1 \neq \theta_2$.

Teorema 3. *Si se cumplen las condiciones* (A_0), (A_c), *y* $0 < I(\theta) \leqslant \leqslant 4h < \infty$ *para todos* $\theta \in \Theta$, *entonces existe una constante* $g > 0$ *tal que para todos* $\theta_1, \theta_2 \in \Theta$,

$$g \leqslant \frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \leqslant h. \tag{9}$$

Demostración. La estimación superior se deduce directamente de (7). Mostremos ahora, que

$$\inf_{\theta_1, \theta_2} \frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \geqslant g > 0. \tag{10}$$

Supongamos que (10) no es cierta, entonces habrá una sucesión $(\theta_1^{(n)}, \theta_2^{(n)}$ tal que

$$\frac{r(\theta_1^{(n)}, \theta_2^{(n)}}{|\theta_1^{(n)} - \theta_2^{(n)}|^2} \to 0 \tag{11}$$

cuando $n \to \infty$. En virtud de la condición (A_c) podemos considerar, sin limitar la generalidad, que $\theta_1^{(n)} \to \theta_1 \in \Theta$, $\theta_2^{(n)} \to \theta_2 \in \Theta$. Si $\theta_1 \neq \theta_2$, entonces (11) contradice (5), ya que, en virtud de la condición (A_0), $r(\theta_1, \theta_2) > 0$. Pero si $\theta_1 = \theta_2 = \theta$, entonces (11) contradice (6), ya que $I(\theta) > 0$. El teorema queda demostrado.

4. Caso multidimensional. En este apartado obtendremos los análogos de las afirmaciones de los puntos 2 y 3 para el parámetro multidimensional (el contenido del punto 1 no está relacionado con la dimensión de θ). Designemos por $\varphi(x, \theta)$ la función vectorial con coordenadas

$$\varphi_i(x, \theta) = \frac{1}{\sqrt{f_\theta(x)}} \frac{\partial f_\theta(x)}{\partial \theta_i}.$$

Entonces la derivada de la función $\sqrt{f_\theta(x)}$ en el sentido del vector unitario $\omega = (\omega_1, \ldots, \omega_k)$ es igual a $((\sqrt{f_\theta(x)})', \omega) = (\text{grad}\sqrt{f_\theta(x)}, \omega) = \frac{1}{2}(\varphi(x, \theta), \omega)$. La matriz de Fisher $I(\theta)$ en estas designaciones es igual a

$$I(\theta) = \int \varphi^T(x, \theta), \varphi(x, \theta)\mu(dx).$$

Supongamos que $|u|$ significa la norma euclídea $u = (u_1, \ldots, u_k)$.

En el caso multidimensional tiene lugar la siguiente generalización del lema 1.

Lema 1A. *La primera afirmación del lema 1* (véase (5)) *conserva por completo su validez cuando* $k > 1$.

Si la función $\sqrt{f_\theta(x)}$, *cuando c.t.* $[\mu]$ *valores de* x, *es derivable respecto a* θ, $\theta' \to \theta$, $\theta'' = \theta' + \omega''\delta$, $\omega'' \to \omega$, $|\omega''| = |\omega| 1$, $\delta \to 0$, *entonces*

$$\text{lím inf} \frac{r(\theta', \theta'')}{|\theta' - \theta''|^2} \geqslant \frac{1}{4} \omega I(\theta) \omega^T. \tag{12}$$

Además, si ω, $|\omega| = 1$ *es un vector colineal a* $\theta_2 - \theta_1$, *de modo que* $\theta_2 = \theta_1 + a\omega$, $a = |\theta_2 - \theta_1|$, *entonces*

$$\frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \leqslant \frac{1}{4} \int_0^1 \omega I(\theta_1 + a\omega y)\omega^T dy. \tag{13}$$

Demostración. La primera afirmación del lema 1 no está relacionada con la dimensión. La segunda se deduce del lema de Fatou y de las relaciones

$$\liminf \frac{r(\theta', \theta'')}{|\theta' - \theta''|^2} \geqslant \int \liminf \frac{(\sqrt{f_{\theta'}} - \sqrt{f_{\theta''}})^2}{|\theta' - \theta''|^2}\, \mu(dx) =$$

$$= \frac{1}{4} \int \varphi((x, \theta), \omega)^2 \mu(dx) = \frac{1}{4}\, \omega I(\theta)\omega^T.$$

Para demostrar (13) indicaremos que

$$\sqrt{f_{\theta_2}} - \sqrt{f_{\theta_1}} = \frac{1}{2} \int_0^a (\varphi(x, \theta_1 + y\omega), \omega)dy =$$

$$= \frac{a}{2} \int_0^1 (\varphi(x, \theta_1 + ay\omega), \omega)dy;$$

$$r(\theta_1, \theta_2) = \frac{a^2}{4} \int_{\mathscr{X}} \left[\int_0^1 (\varphi(x, \theta_1 + ay\omega), \omega dy \right]^2 \mu(dx) \leqslant$$

$$\leqslant \frac{a^2}{4} \int_{\mathscr{X}} \int_0^1 (\varphi(x, \theta_1 + ay\omega), \omega)^2 dy \mu(dx) =$$

$$= \frac{a^2}{4} \int_0^1 \int_{\mathscr{X}} (\varphi(x, \theta_1 + ay\omega), \omega)^2 \mu(dx)dy = \frac{a^2}{4} \int_0^1 \omega I(\theta_1 + ay\omega)\omega^T dy. \quad \triangleleft$$

Pongamos, como antes, $r(\Delta) = r(\theta, \theta + \Delta)$. Del lema 1A se deduce el

Teorema 2A. *Si la función* $\sqrt{f_\theta(x)}$ *es derivable cuando c.t.* $|\mu|$ *valores de* x, *y la matriz* $I(\theta)$ *es continua, entonces para cualquier vector* ω *de longitud unitaria existe*

$$\lim_{\delta \to 0} \frac{r(\delta\omega)}{\delta^2} = \frac{1}{4}\, \omega I(\theta)\omega^T.$$

Al igual que en el caso unidimensional, del lema 1*A* también podemos obtener el corolario siguiente. Designemos por Sp $I(\theta)$ la traza de la matriz $I(\theta)$.

Teorema 3A. *Si se cumplen las condiciones* (A_0), (A_c), *y la matriz* $I(\theta)$ *es positivamente definida en* Θ, $4h = \sup \operatorname{Sp} I(\theta) < \infty$, *entonces existe una constante* $g > 0$ *tal que para todos* $\theta_1, \theta_2 \in \Theta$

$$g \leqslant \frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \leqslant h. \tag{14}$$

Demostración. Designemos por $\Lambda_1(\theta)$ y $\Lambda_k(\theta)$ los números propios, mínimo y máximo, respectivamente, de la matriz $I(\theta)$, así que cuando $|\omega| = 1$,

$$\Lambda_1(\theta) \leqslant \omega I(\theta)\omega^T \leqslant \Lambda_k(\theta). \tag{15}$$

Según las condiciones del teorema, $\Lambda_1(\theta) > 0$ siempre en Θ. Como $(\varphi, \omega)^2 \leqslant |\varphi|^2 = \sum_{j=1}^{k} \varphi_j^2$, entonces

$$\int_{\mathscr{X}} (\varphi, \omega)^2 \mu(dx) = \omega I(\theta)\omega^T \leqslant \operatorname{Sp} I(\theta)$$

y, por consiguiente, $\Lambda_k(\theta) \leqslant \operatorname{Sp} I(\theta) \leqslant 4h$. De la desigualdad (13) obtenemos

$$\frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \leqslant \frac{1}{4} \int_0^1 \Lambda_k(\theta_1 + ay\omega)dy \leqslant h.$$

Demostremos ahora la segunda desigualdad en (14). Supongamos que ésta no es cierta. Entonces, al igual que en el teorema 3, habrá una sucesión $(\theta_1^{(n)}, \theta_2^{(n)})$, $\theta_1^{(n)} \to \theta_1 \in \Theta$, $\theta_2^{(n)} \to \theta_2 \in \Theta$, para la cual será válida (11). Si $\theta_1 \neq \theta_2$, esto contradirá (5). Si $\theta_1 = \theta_2 = \theta$, entonces, en virtud de la compacticidad de la esfera $|\omega| = 1$, se puede considerar, sin limitar la generalidad, que $\theta_2^{(n)} = \theta_1^{(n)} + \delta\omega^{(n)}$, $\omega^{(n)} \to \omega$, $|\omega^{(n)}| = |\omega| = 1$. Pero en este caso (11) contradirá (12) y (15). ◁

5.* Relación entre las distancias sujetas a examen y las estimaciones. Examinemos la distancia de Kullback—Leibler entre la distribución $\mathbf{P}_\theta$ y la distribución $\mathbf{G}$ que no depende de θ:

$$\varrho_1(\mathbf{G}, \mathbf{P}_\theta) = \int \ln \frac{d\mathbf{G}}{d\mu} \mathbf{G}(dx) - \int \ln f_\theta(x) \mathbf{G}(dx).$$

Aquí sólo depende de θ el segundo sumando

$$d(\mathbf{P}_\theta, \mathbf{G}) = -\int \ln f_\theta(x) \mathbf{G}(dx).$$

Por otro lado, recordemos que la e.v.m. ha sido definida en el § 6 como valor de θ con el que se minimiza $d(\mathbf{P}_\theta, \mathbf{P}_n^*)$. Si la distribución de x_1 es discreta, y μ es la medida de cálculo, entonces la expresión

$$d(\mathbf{P}_n^*, \mathbf{P}_n^*) = -\int \ln \frac{d\mathbf{P}_n^*}{d\mu} \mathbf{P}_n^*(dx)$$

tiene sentido, $\varrho_1(\mathbf{P}_n^*, \mathbf{P}_\theta) = d(\mathbf{P}_\theta, \mathbf{P}_n^*) - d(\mathbf{P}_n^*, \mathbf{P}_n^*)$ y, por consiguiente, podemos considerar que la e.v.m. minimiza la distancia de Kullback—Leibler

$\varrho_1(\mathbf{P}_n^*, \mathbf{P}_\theta)$ entre $\mathbf{P}_\theta$ y $\mathbf{P}_n^*$. En el caso general tal interpretación puede ser aceptada sólo convencionalmente.

Para las distribuciones discretas de x_1 también se pueden examinar las distancias $\varrho_i(\mathbf{P}_\theta, \mathbf{P}_n^*)$ cuando $i = 2, 3$, así como las estimaciones que minimizan estas distancias. Por ejemplo, cuando $i = 2$ obtenemos

$$\varrho_2(\mathbf{P}_\theta, \mathbf{P}_n^*) = \sum_i \frac{\left(\frac{\nu_i}{n} - f_\theta(a_i)\right)^2}{f_\theta(a_i)},$$

donde ν_i es el número de elementos de la muestra, los cuales han caído en el punto a_i, para el cual $f_\theta(a_i) = \mathbf{P}_\theta(\{a_i\} > 0$. Esta es la estadística χ^2 (véase los §§ 7 y 8), debido a lo cual también hemos dado tal denominación a la distancia ϱ_2.

En vista de que las distancias ϱ_i poseen propiedades asintóticas semejantes, las estimaciones que las minimizan, como será aclarado más tarde, coincidirán asintóticamente.

§ 22.* Desigualdad de diferencias del tipo Rao—Cramer

Este párrafo está un poco apartado de la exposición principal. Aquí trataremos de responder, aunque sea parcialmente, a la pregunta acerca de qué es lo que ocurre con la frontera inferior admisible para $\mathbf{M}_\theta(\theta^* - \theta)^2$ en el caso irregular, o sea, en el caso cuando la función $f_\theta(x)$ no es derivable respecto a θ o cuando $I(\theta) = \infty$.

Comenzaremos por el ejemplo que muestra que, en estas condiciones, el comportamiento de las desviaciones estándar (o de sus varianzas) puede diferenciarse totalmente del segundo miembro de la desigualdad de Rao—Cramer.

Ejemplo 1. Sea $X \in \mathbf{U}_{0\theta}$. Aquí, la condición ($R$) no se cumple, ya que la función $f_\theta(x)$ es discontinua. Como sabemos, para esta familia estadística $S = \max x_i$ es completa y suficiente (véase el ejemplo 14.3). Tomemos la estimación no desplazada $\theta^* = 2x_1$. Entonces, en virtud de los resultados obtenidos en el § 14, la estadística $\theta_S^* = 2\mathbf{M}_\theta(x_1/S)$ será eficiente. Calculemos el valor de $\mathbf{M}_\theta(x_1/S)$. Como $\mathbf{P}_\theta(S < z) = (z/\theta)^n$, $z \in [0, \theta]$, entonces S tiene una densidad igual a nz^{n-1}/θ^n en $[0, \theta]$ e igual a cero fuera de ese intervalo. Para hallar la distribución condicional $P(B/s) = \mathbf{P}_\theta(x_1 \in B/S) = s)$ de la magnitud x_1, a condición de que $S = s$, utilizaremos la regla (10.2):

$$P(dy/s) = \mathbf{P}_\theta(x_1 \in dy/S = s) = \frac{\mathbf{P}_\theta(x_1 \in dy,\ S \in ds)}{\mathbf{P}_\theta(S \in dx)}.$$

Aquí el numerador es igual a

$$\mathbf{P}_\theta(x_1 \in dy,\ S \in dx) = \begin{cases} \dfrac{dy}{\theta} \cdot \dfrac{(n-1)s^{n-2}ds}{\theta^{n-1}} & \text{cuando } y < s, \\ \dfrac{ds}{\theta} \cdot \dfrac{s^{n-1}}{\theta^{n-1}} & \text{cuando } y = s, \\ 0 & \text{cuando } y > s. \end{cases}$$

De aquí se deduce que $P(dy/s) = \dfrac{(n-1)dy}{ns}$ cuando $0 \leqslant y < s$, $P(\{s\}/s) = 1/n$. Por lo tanto,

$$\mathbf{M}_\theta x_1/S) = \int_0^\theta y \frac{n-1}{nS}\,dy + \frac{S}{n} = \frac{S(n-1)}{2n} + \frac{S}{n} = \frac{n+1}{2n}S,$$

$$\theta_S^* = S\left(1 + \frac{1}{n}\right).$$

Tenemos

$$\mathbf{D}_\theta\theta_S^* = \mathbf{M}_\theta(\theta_S^*)^2 - \theta^2 = \int_0^\theta s^2\left(1 + \frac{1}{n}\right)^2 \frac{ns^{n-1}}{\theta^n}\,ds - \theta^2 =$$

$$= \left(\frac{(n+1)^2}{n(n+2)} - 1\right)\theta^2 = \frac{\theta^2}{n(n+2)}. \quad (1)$$

Como θ_S^* es eficiente, para toda estimación no desplazada θ^*,

$$\mathbf{D}_\theta\theta^* \geqslant \frac{\theta^2}{n(n+2)}. \quad (2)$$

Ahora bien, para grandes valores de n, la desviación estándar de $\mathbf{M}_\theta(\theta_S^* - \theta)^2$ tendrá un orden de pequeñez de $1/n^2$. Desde el punto de vista de la frontera inferior de la desigualdad de Rao—Cramer, que tiene un orden de $1/n$, la misma constituye una exactitud anormalmente alta*). Se puede mostrar que ésta es la exactitud con la que, a partir de la muestra, se determinan cualesquiera puntos de saltos de $f_\theta(x)$ prohibidos por la condición (R)). En el ejemplo 7.4, dedicado a la estimación de la mediana, hemos visto que los puntos donde la densidad $f_\theta(x)$ es infinita, se pueden determinar aún más exactamente, así que, en términos generales, cuanto mayor sea la alteración de la regularidad en el punto, tanto más exactamente será apreciado este punto por la muestra. Digamos, si $X \in \mathbf{P}_\theta$, donde $\mathbf{P}_\theta = \frac{1}{2}\mathbf{U}_{0,\theta} + \frac{1}{2}\mathbf{I}_\theta$, $\mathbf{I}_\theta$ es la distribución concentrada en el punto θ, entonces

---

*) Para el parámetro θ también existen estimaciones cuya varianza tiene el orden de $1/n$. Por ejemplo, para la estimación $\theta^{**} = 2\bar{x}$ tenemos $\mathbf{M}\theta^{**} = \theta$, $\mathbf{D}\theta^{**} = \frac{4}{n}Dx_1 = \frac{\theta^2}{3n}$.

$\mathbf{P}_\theta(S \neq \theta) = 2^{-n}(S = \text{máx}\, x_i)$, así que la varianza de $\theta^* - \theta$, cuando $\theta^* = S$, decrecerá exponencialmente con el aumento de n.

¿Será posible en estas condiciones indicar la frontera inferior para la varianza de las estimaciones? Más adelante obtendremos una desigualdad análoga a la de Rao—Cramer, mediante la cual tales fronteras pueden ser construidas cuando las condiciones de regularidad son menos rigurosas que la condición (R).

Solamente supondremos que se cumple la condición (A_μ), aunque tampoco eso tiene mucha importancia (véase la observación al final del párrafo).

Designemos por $\Delta\varphi(\theta)$ el incremento de la función $\varphi(\theta)$ en el intervalo $(\theta, \theta + \Delta)$; por $N^n_{P_\theta}$, el portador en $\mathscr{X}^n$ de la distribución de la muestra: $N^n_{P_\theta} = \{x: f_\theta(x) \neq 0\}$ y pongamos $N^n = N^n_{P_\theta} \cup N^n_{P_{\theta+\Delta}}$.

Teorema 1. (Desigualdad de Chapman—Robbins). *Sea* $\theta \in \Theta$, $\theta + \Delta \in \Theta$, $a(\theta) = \mathbf{M}_\theta\theta^*$. *Entonces, para cualquier* $\Delta \neq 0$,

$$\mathbf{D}_\theta\theta^* \geqslant \frac{(\Delta_a(\theta))^2}{\int [\Delta f_\theta(x)]^2/f_\theta(x)\mu^n(dx)} = \frac{(\Delta a(\theta))^2}{\varrho_2(\mathbf{P}^n_{\theta+\Delta}, \mathbf{P}^n_\theta)}, \tag{3}$$

donde ϱ_2 *es la distancia* χ^2 *examinada en el § 21. Aquí, para las estimaciones no desplazadas es necesario sustituir el numerador por* Δ^2.

En virtud del teorema 21.1, el denominador en (3) tiene la forma $\varrho_2(\mathbf{P}^n_{\theta+\Delta}, \mathbf{P}^n_\theta) = (1 + r_2(\Delta))^n - 1$, donde

$$r_2(\Delta) = \varrho_2(\mathbf{P}_{\theta+\Delta}, \mathbf{P}_\theta) = \int \frac{[\Delta f_\theta(x)]^2}{f_\theta(x)} \mu(dx). \tag{4}$$

Ahora bien, cuanto mayor sea la distancia $\varrho_2(\mathbf{P}_{\theta+\Delta}, \mathbf{P}_\theta)$ entre $\mathbf{P}_{\theta+\Delta}$ y $\mathbf{P}_\theta$ (al ser registrado Δ), tanto menor será la frontera inferior para $\mathbf{D}\theta^*$.

Si $\mathbf{P}_{\theta+\Delta}$ es absolutamente continua respecto a $\mathbf{P}_\theta$ entonces $N^n_{P_{\theta+\Delta}} \subset N^n_{P_\theta} = N^n$ $\varrho_2(\mathbf{P}^n_{\theta+\Delta}, \mathbf{P}^n_\theta)$ puede escribirse en la forma (véase (21.2))

$$\varrho_2(\mathbf{P}^n_{\theta+\Delta}, \mathbf{P}^n_\theta) = \mathbf{M}_\theta\left[\frac{\Delta f_\theta(X)}{f_\theta(X)}\right]^2;$$

análogamente, $r_2(\Delta) = \mathbf{M}_\theta\left[\dfrac{\Delta f_\theta(x_1)}{f_\theta(x_1)}\right]^2$.

Pero si la distribución $\mathbf{P}_{\theta+\Delta}$ no es absolutamente continua respecto a $\mathbf{P}_\theta$, entonces existe un subconjunto de $N_{P_{\theta+\Delta}}$ de medida positiva $\mathbf{P}_{\theta+\Delta}$, en el que $f_\theta(x) = 0$, así que la integral en (4) se vuelve infinita, y la propia desigualdad (3) se vuelve trivial. Es necesario señalar otra vez, que en este caso la expresión $\mathbf{M}_\theta[\Delta f_\theta(X)/f_\theta(X)]^2$, entendida como integral respecto a N_{P_θ}, puede permanecer finita.

Demostración del teorema 1. De lo dicho anteriormente se deduce que, sin limitar la generalidad, podemos considerar que $\mathbf{P}_{\theta+\Delta}$ es absolutamente continua respecto a $\mathbf{P}_\theta$, así que $N^n_{\mathbf{P}_{\theta+\Delta}} \subset N^n_{\mathbf{P}_\theta} = N^n$. Como $f_\theta(x)$ y $f_{\theta+\Delta}(x)$ es la densidad en $\mathscr{X}^n$, entonces

$$\int \Delta f_\theta(x)\mu^n(dx) = 0.$$

Además,

$$\int \theta^* \Delta f_\theta(x)\mu^n(dx) = \Delta a(\theta).$$

De aquí se desprende que

$$\int_{N^n} (\theta^* - a(\theta))\Delta f_\theta(x)\mu^n(dx) = \Delta a(\theta). \qquad (5)$$

En el conjunto N^n podemos representar la función subintegral de (5) en forma del producto

$$(\theta^* - a(\theta))^2\sqrt{f_\theta(x)} \cdot \frac{\Delta f_\theta(x)}{\sqrt{f_\theta(x)}}.$$

Aplicando luego la desigualdad de Cauchy—Buniakovski, obtenemos

$$(\Delta a(\theta))^2 \leqslant \int_{N^n} (\theta^* - \theta)^2 f_\theta(x)\mu^n(dx) \int_{N^n} \frac{(\Delta f_\theta(x))^2}{f_\theta(x)}\,\mu^n(dx). \quad \triangleleft$$

En lo sucesivo, según las observaciones hechas más arriba, nos limitaremos, al igual que en la demostración del teorema 1, al caso cuando $\mathbf{P}_{\theta+\Delta}$ es absolutamente continua respecto a $\mathbf{P}_\theta$ (de lo contrario la desigualdad (3) se vuelve trivial).

Corolario 1. *Si se cumplen las condiciones de regularidad que aseguran la existencia* (véase la observación 21.1 al teorema 21.2) de $\lim_{\Delta\to 0} r_2(\Delta)/\Delta^2 = I(\theta)$, *entonces*

$$\mathbf{D}_\theta\theta^* \geqslant \frac{(a'_+(\theta))^2}{nI(\theta)}, \qquad (6)$$

donde $a'_+(\theta) = \lim_{\Delta\to 0}\sup \dfrac{\Delta a(\theta)}{\Delta}$.

Para obtener (6) del teorema 1 sólo es necesario notar que podemos elegir la sucesión $\Delta \to 0$ de modo que $\dfrac{\Delta a(\theta)}{\Delta} \to a'_+(\theta)$. $\triangleleft$

La desigualdad (6) es, según su forma, cierta generalización de la desigualdad de Rao—Cramer (generalización, lo más probable, ficticia, ya que las condiciones de regularidad mencionadas conducen, por lo visto, a la existencia de $a'(\theta)$).

La desigualdad (3), por supuesto, se denomina desigualdad de *diferencias*, a distinción de la desigualdad (6) que podría denominarse desigualdad *diferencial*.

Ahora bien, si $r_2(\Delta) \sim I(\theta)\Delta^2$ (esto corresponde al hecho de que f_θ es derivable), *entonces de la desigualdad de diferencias de Chapman—Robbins se deduce la desigualdad diferencial de Rao—Cramer.*

Pero si la función f_θ no es derivable, entonces, al disminuir Δ, el comportamiento de $r_2(\Delta)$ será diferente.

Si, digamos, f_θ es derivable en todas partes, a excepción de un número finito de puntos de discontinuidad $\theta = \theta(x)$ que dependen de x, entonces tendremos

$$r_2(\Delta) \sim c|\Delta|. \tag{7}$$

Esto puede ser aclarado de la forma más sencilla a base de un ejemplo muy típico, examinado al principio del párrafo.

Sea $X \in \mathrm{U}_{0,\theta}$. Para que sea cumplida la condición de continuidad absoluta de $\mathbf{P}_{\theta+\Delta}$ respecto a $\mathbf{P}_\theta$, en el caso de $\mathbf{P}_\theta = \mathrm{U}_{0,\theta}$ consideraremos que $\Delta < 0$, $|\Delta| < \theta$. Entonces

$$\Delta f_\theta(x) = \begin{cases} \dfrac{1}{\theta+\Delta} - \dfrac{1}{\theta} & \text{para } x \in [0, \theta+\Delta], \\ -\dfrac{1}{\theta} & \text{para } x \in [0+\Delta, \theta], \\ 0 & \text{para } x \notin [0, \theta], \end{cases}$$

$$r_2(\Delta) = \int_0^\theta \frac{(\Delta f_\theta(x))^2}{f_\theta(x)}\,dx = \int_0^{\theta+\Delta} \left[\frac{\Delta}{\theta(\theta+\Delta)}\right]^2 \theta\,dx = \int_{\theta+\Delta}^{\theta} \frac{1}{\theta^2}\,\theta\,dx =$$

$$= \frac{\Delta^2}{\theta(\theta+\Delta)} + \frac{|\Delta|}{\theta}.$$

Lo esencial aquí es la existencia del intervalo cuya longitud es comparable con Δ y en el que $|\Delta f_\theta(x)| > c > 0$, donde c no depende de Δ. Esto asegura precisamente el orden de pequeñez (7) para $r_2(\Delta)$.

Volviendo a nuestro ejemplo, vemos que para las estimaciones no desplazadas del parámetro θ,

$$\mathbf{D}\theta^* \geqslant \max_\Delta \frac{\Delta^2}{\left(1 + \dfrac{|\Delta|}{\theta} + \dfrac{\Delta^2}{\theta(\theta+\Delta)}\right)^n - 1}.$$

¿Cuál es el orden de pequeñez del segundo miembro de esta desigualdad cuando $n \to \infty$? Suponiendo $|\Delta| = y\theta/n$, obtenemos

$$\mathbf{D}\theta^* \geqslant \frac{\theta^2}{n^2} \max_y \frac{y^2}{\left(1 + \frac{y}{n} + \frac{y^2}{n(n-y)}\right)^n - 1}.$$

Está claro que la expresión con signo máx es asintóticamente equivalente a $h = \max y^2/(e^y - 1) \approx 0{,}65$, así que

$$\mathbf{D}\theta^* \geqslant \frac{\theta^2}{n^2}(h + o(1)).$$

En cuanto al orden de pequeñez, esta desigualdad tiene el mismo segundo miembro que la desigualdad inmejorable (2), pero el factor constante de θ^2/n^2 en (2) es "mejor" y es igual a 1.

A la par con (7) pueden aparecer también otras velocidades de convergencia de $r_2(\Delta)$ hacia el cero, cuando $\Delta \to 0$. Podemos obtener, por ejemplo, tanto $r_2(\Delta) \sim c\Delta^\alpha$, $\alpha < 1$, si $f_\theta(x)$ tiene líneas de $\theta = \theta(x) \neq$ const, al aproximarse a las cuales $f_\theta(x) \to \infty$; como también $r_2(\Delta) \sim c\Delta^\alpha$, $2 > \alpha > 1$, si f_θ es continua respecto a θ pero no es derivable sino satisface solamente la condición de Hölder en el entorno de cierta línea $\theta = \theta(x) \neq$ const. No es difícil ver que el orden de pequeñez

$$\max_\Delta \frac{\Delta^2}{(1 + c\Delta^\alpha)^n - 1}$$

para $\alpha < 2$ será definido por el valor de $\Delta = (y/cn)^{1/\alpha}$, así que

$$\mathbf{D}\theta^* \geqslant \frac{1}{(cn)^{2/\alpha}} \max_y \frac{y^{2/\alpha}}{e^y - 1}(1 + o(1)).$$

En el caso "regular" $\alpha = 2$, el máximo respecto a y se obtiene en el punto límite $y = 0$ $(\Delta = 0)$.

Concluyendo este párrafo señalaremos que las estimaciones para $\mathbf{D}\theta^*$ también pueden ser obtenidas, de modo análogo, para las no absolutamente bicontinuas $\mathbf{P}_\theta$ y $\mathbf{P}_{\theta+\Delta}$. Para esto, en (5) es necesario multiplicar y dividir la función subintegral no por $\sqrt{f_\theta(x)}$, sino por $\sqrt{f_\theta(x) + f_{\theta+\Delta}(x)}$. La condición (A_μ) tampoco es tan esencial, ya que las medidas de $\mathbf{P}_{\theta_1}$ y $\mathbf{P}_{\theta+\Delta}$ siempre son absolutamente continuas respecto a $\frac{1}{2}(\mathbf{P}_\theta + \mathbf{P}_{\theta+\Delta})$.

§ 23. Desigualdades auxiliares para la relación de verosimilitud. Conciliabilidad de las estimaciones de la verosimilitud máxima

En los §§ 12—16 hemos estudiado las cuestiones relacionadas con la existencia y la determinación, en forma explícita, de las estimaciones eficientes y R-eficientes. Hemos visto que éstas existen no siempre, ni mucho menos, y pueden ser halladas tan sólo en el caso cuando la función de verosimilitud tiene una forma especial o cuando conocemos, de manera explícita, la estadística suficiente completa (la primera de estas condiciones a menudo conduce a la segunda (véase el § 15)).

Pasemos ahora a la construcción de las estimaciones asintóticamente óptimas. Aquí las condiciones de su existencia serán mucho más amplias. Los resultados respectivos se apoyan, ante todo, en las propiedades asintóticas de la función

$$Z(u) = \frac{f_{\theta+u}(X)}{f_\theta(X)} = \exp\{L(X, \theta+u) - L(X, \theta)\}, \qquad (1)$$

donde, como antes, $L(X, \theta) = \sum_{i=1}^{n} l(x_i, \theta)$. Por regla general, el número θ en (1) se considerará registrado y representará el valor real del parámetro, o sea, tal que $X \in \mathbf{P}_\theta$. En este caso $Z(u)$ es la función de los variables u y X y, por lo tanto, junto con la función de verosimilitud $f_{\theta+u}(X)$, será la *función aleatoria* de la variable u. Llamaremos *relación de verosimilitud* la función $Z(u)$ que desempeña un papel muy importante en la estadística matemática. La tarea principal de este párrafo y del párrafo siguiente consiste en estudiar las propiedades de $Z(u)$.

Será establecido que $Z(u)$ es próxima a cero fuera del entorno del punto $u = 0$. En el entorno de este punto, $Z(u)$ se aproxima, desde cierto punto de vista, a la función delta, mejor dicho, $Z(v/\sqrt{n})$ se aproxima asintóticamente, cuando $n \to \infty$, a la función de densidad de la ley normal.

En los §§ 23—26 examinaremos sólo el parámetro unidimensional. El caso del parámetro multidimensional será investigado separadamente en el § 28.

En las estimaciones posteriores desempeñará un gran papel la distancia de Hellinger

$$r(u) = \varrho(\mathbf{P}_{\theta+u}, \mathbf{P}_\theta) = \int \left(\sqrt{f_{\theta+u}(x)} - \sqrt{f_\theta(x)}\right)^2 \mu(dx)$$

entre las distribuciones $\mathbf{P}_{\theta+u}$ y $\mathbf{P}_\theta$. Hemos examinado esta distancia en el § 21. Recordemos que

$$0 \leqslant r(u) = 2\left(1 - \int \sqrt{f_{\theta+u}(x) f_\theta(x)}\, \mu(dx)\right) \leqslant 2,$$

así que

$$\mathbf{M}_\theta \sqrt{\frac{f_{\theta+u}(\mathrm{x}_1)}{f_\theta(\mathrm{x}_1)}} = \int \sqrt{f_{\theta+u}(x) f_\theta(x)}\, \mu(dx) = 1 - r(u)/2, \qquad (2)$$

$$\mathbf{M}_\theta Z^{1/2}(u) = (1 - r(u)/2)^n. \qquad (3)$$

En lo que se refiere a la familia paramétrica $\{\mathbf{P}_\theta\}$, supondremos en este párrafo y en los párrafos siguientes que *a la par con* (A_μ) *se cumplen las condiciones* (A_0) $(f_{\theta_1}(x) \neq f_{\theta_2}(x)$ *para* $\theta_1 \neq \theta_2$ y (A_c) (Θ *es un compacto*). El hecho de que la última condición es poco importante desde el punto de vista de aplicaciones, ha sido mencionado anteriormente. Esto se debe a que en los problemas reales, de ordinario es posible señalar las fronteras de los posibles valores de θ, partiendo de las consideraciones a priori. Para simplificar la exposición, allí donde sea necesario, también supondremos que Θ es convexo (en el caso unidimensional esto quiere decir que $\Theta = [a, b]$, $-\infty < a < b < \infty$).

Además, en este párrafo supondremos que la función $\sqrt{f_\theta}$ es *derivable para c.t.* $[\mu]$ *valores de x, y que la información de Fisher*

$$I(\theta) = \int \frac{(f'_\theta(x))^2}{f_\theta(x)} \mu(dx) = \mathbf{M}_\theta \left(\frac{f'_\theta(\mathrm{x}_1)}{f_\theta(\mathrm{x}_1)}\right)^2$$

es estrictamente positiva y está limitada en Θ. En estas condiciones hemos demostrado en el teorema 21.3 que para todos θ y $\theta + u$ admisibles (o sea, tales que $\theta \in \Theta$, $\theta + u \in \Theta$) para la magnitud $r(u) = \varrho(\mathbf{P}_{\theta+u}, \mathbf{P}_\theta)$ es válida la desigualdad

$$\inf_{\theta, u} \frac{r(u)}{u^2} \geqslant g > 0. \qquad (4)$$

1. **Desigualdades principales.** Designemos, para abreviar, $p(u) = Z^{3/4}(u)$ y supongamos que se cumplen todas las condiciones anteriormente citadas.

Teorema 1.

$$\mathbf{M}_\theta Z^{1/2}(u) \leqslant e^{-ngu^2/2}, \quad \mathbf{M}_\theta p(u) \leqslant e^{-ngu^2/4}, \qquad (5)$$

$$\mathbf{M}_\theta |p'(u)| \leqslant \frac{3}{4} \sqrt{nI(\theta + u)}\, e^{-u^2 ng/4}.$$

De las investigaciones realizadas en el § 21 se deduce que para los valores $u = o(1)$ en estas desigualdades, en vez de g se pueden tomar los valores tan próximos como se quiera a $I(\theta)$.

Demostración. En virtud de (3) y (4) tenemos

$$\mathbf{M}_\theta Z^{1/2}(u) = (1 - r(u)/2)^n \leqslant \exp\{-nr(u)/2\} \leqslant \exp\{-ngu^2/2\}.$$

Luego, en virtud de la desigualdad de Cauchy — Buniakovski,

$$\mathbf{M}_\theta p(u) \leqslant [\mathbf{M}_\theta Z^{1/2}(u)\cdot \mathbf{M}_\theta Z(u)]^{1/2} = [\mathbf{M}_\theta Z^{1/2}(u)]^{1/2} \leqslant e^{-u^2ng/4}.$$

Volviendo a utilizar la desigualdad de Cauchy — Buniakovski y la relación

$$p'(u) = \frac{3}{4} L'(X, \theta + u) Z^{3/4}(u),$$

hallamos

$$\mathbf{M}_\theta |p'(u)| = \frac{3}{4}\mathbf{M}_\theta |L'(X, \theta +)| Z^{1/2}(u) Z^{1/4}(u) \leqslant$$

$$\leqslant \frac{3}{4} [\mathbf{M}_\theta [L'(X, \theta + u)]^2 Z(u)\cdot \mathbf{M}_\theta Z^{1/2}(u)]^{1/2} \leqslant$$

$$\leqslant \frac{3}{4} [\mathbf{M}_{\theta+u} [L'(X, \theta + u)]^2]^{1/2} e^{-u^2ng/4}. \quad \triangleleft$$

Teorema 2. *Para todos z, $n \geqslant 1$*

$$\mathbf{P}_\theta\Big(\sup_{|v|\geqslant u} Z(v/\sqrt{n})\Big) \leqslant ce^{-3z/4} e^{-u^2g/4},$$

donde $c = 2 + 3\sqrt{\pi I_0/g}$, $I_0 = \sup_{\theta\in\Theta} I(\theta)$ no dependen de θ.

Para demostrar el teorema necesitaremos el

Lema 1. *Para todos $x \geqslant 0$,*

$$\int_x^\infty e^{-v^2/2} dv \leqslant \sqrt{2\pi}\, e^{-x^2/2}.$$

Demostración*). La función característica de la variable aleatoria $\xi \in \Phi_{0,1}$ es igual a $\mathbf{M}e^{it\xi} = e^{-t^2/2}$ y está definida en todo el plano. Suponiendo $t = -ix$, obtendremos $\mathbf{M}e^{x\xi} = e^{x^2/2}$. De aquí, con ayuda de la desigualdad de Chébishev, obtenemos

$$\mathbf{P}(\xi > x) = \mathbf{P}(e^{\xi x} > e^{x^2}) \leqslant e^{-x^2}\mathbf{M}e^{\xi x} = e^{-x^2/2}. \quad \triangleleft$$

Demostración del teorema 2. Estimemos la función

$$H(\delta) = \mathbf{M}_\theta \sup_{|v|>\delta} p(v).$$

Si $v \in [\theta + \delta, b]$, entonces

$$p(v - \theta) = p(\delta) + \int_\delta^{v-\theta} p'(u)du \leqslant p(\delta) + \int_\delta^{b-\theta} |p'(u)|du.$$

---

*) Para grandes x son más exactas las desigualdades siguientes:

$$\frac{1}{x+1} e^{-x^2/2} < \int_x^\infty e^{-v^2/2} dv < \frac{1}{x} e^{-x^2/2},$$

las cuales pueden ser fácilmente obtenidas por el lector, comparando las derivadas de las funciones sujetas a examen (los valores de las propias funciones coinciden cuando $x = \infty$).

Como aquí el segundo miembro no depende de v, entonces

$$\sup_{u\geqslant\delta} p(u) \leqslant p(\delta) + \int_{u\geqslant\delta} |p'(u)|du,$$

$$H_+(\delta) \equiv \mathbf{M}_\theta \sup_{u\geqslant\delta} p(u) \leqslant \mathbf{M}_\theta p(\delta) + \int_{u\geqslant\delta} \mathbf{M}_\theta|p'(u)|du.$$

De aquí, en virtud del teorema 1 obtenemos

$$H_+(\delta) \leqslant e^{-\delta^2 ng/4} + \frac{3}{4}\sqrt{n} \int_{u\geqslant\delta} \sqrt{I(\theta+u)}\, e^{-u^2 ng/4}du.$$

A base del lema 1,

$$H_+(\delta) \leqslant e^{-ng\delta/4} + \frac{3}{4}\sqrt{nI_0} \int_{|u|\geqslant\delta} e^{-ngu^2/4}du \leqslant$$

$$\leqslant e^{-ng\delta^2/4} + \frac{3}{4}\sqrt{2I_0/g} \int_{v\geqslant\delta\sqrt{ng/2}} e^{-v^2/2}dv \leqslant e^{-ng\delta^2/4}\left(1 + \frac{3}{2}\sqrt{\pi I_0/g}\right).$$

Está claro que una estimación exactamente igual, será válida para la función

$$H_-(\delta) = \sup_{u\leqslant-\delta} p(u).$$

Por eso

$$H(\delta) \leqslant H_+(\delta) + H_-(\delta) \leqslant (2 + 3\sqrt{\pi I_0/g})e^{-ng\delta^2/4}.$$

Queda hacer uso de la desigualdad de Chébishev:

$$\mathbf{P}_\theta(\sup_{|t|\geqslant\delta} Z(t) > e^z) = \mathbf{P}_\theta(\sup_{|t|\geqslant\delta} p(t) > e^{3z/4}) \leqslant H(\delta)e^{-3z/4}. \quad \triangleleft$$

2. Estimaciones para la distribución y los momentos de la e.v.m. Conciliabilidad de la e.v.m.

Teorema 3. *Existen valores de $c<\infty$, $g>0$ tales, que*

$$\mathbf{P}_\theta(\sqrt{n}(\hat{\theta}^* - \theta) \geqslant v) \leqslant ce^{-gv^2/4} \qquad (6)$$

para todos v y $n\geqslant 1$.

Demostración. Del teorema 2 se desprende que

$$\mathbf{P}_\theta(\sup_{|t|\geqslant v/\sqrt{n}} Z(t) > 1) \leqslant ce^{-gv^2/4}.$$

Queda hacer uso de la relación

$$\{|\hat{\theta}^* - \theta| \geqslant \delta\} = \left\{\sup_{|t|\geqslant\delta} Z(t) \geqslant \sup_{|t|\leqslant\delta} Z(t)\right\} \subset \left\{\sup_{|t|\geqslant\delta} Z(t) \geqslant Z(0) = 1\right\} \quad (7)$$

cuando $\delta = v\sqrt{n}$. $\triangleleft$

Corolario 1. *Supongamos que $u_n \to \infty$ es toda sucesión indefinidamente creciente. Entonces*

$$(\hat{\theta}^* - \theta)\sqrt{n}/u_n \underset{P}{\to} 0. \tag{8}$$

No obstante, si u_n son tales que para cualquier $\alpha > 0$

$$\sum e^{-\alpha u_n^2} < \infty, \tag{9}$$

entonces

$$(\hat{\theta}^* - \theta)\sqrt{n}/u_n \underset{c.s.}{\to} 0. \tag{10}$$

Estas relaciones son, evidentemente, las amplificaciones de la conciliabilidad ($\hat{\theta}^* - \theta \underset{P}{\to} 0$) y de la conciliabilidad fuerte ($\hat{\theta}^* - \theta \underset{c.s.}{\to} 0$) de la e.v.m., respectivamente.

Demostración. La relación (8) se deduce directamente de (6) si en esta última se pone $v = \delta u_n$. La relación (10) también se desprende de (6), ya que la suma de los segundos miembros en (6), al cumplirse (9), formará una serie convergente. ◁

Por ejemplo, incluso una sucesión tan lentamente creciente como $u_n = \ln n$ satisface la condición (9), así que*)

$$(\hat{\theta}^* - \theta)\sqrt{n}/\ln n \underset{c.s.}{\to} 0.$$

Corolario 2. *Existe un valor $c_1 < \infty$, no dependiente de n y θ, tal que para todo $\alpha \leqslant g/5$,*

$$\mathbf{M}_\theta \exp\{\alpha(u^*)^2\} < c_1, \textit{ donde } u^* = \sqrt{n}(\hat{\theta}^* - \theta). \tag{11}$$

Demostración. Integrando por partes, obtenemos

$$\mathbf{M}e^{\alpha\xi^2} = -\int_0^\infty e^{\alpha v^2} d\mathbf{P}(|\xi| \geqslant v) = 1 + 2\alpha\int_0^\infty ve^{\alpha v^2}\mathbf{P}(|\xi| \geqslant v)dv.$$

Por eso, en virtud del teorema 3,

$$\mathbf{M}_\theta e^{\alpha(u^*)^2} \leqslant 1 + \frac{2g}{5}\int_0^\infty ve^{-gv^2/20}dv \equiv c_1 < \infty. \quad ◁$$

§ 24. Propiedades asintóticas de la relación de verosimilitud

En el párrafo precedente hemos establecido una serie de desigualdades para $Z(u)$. Determinemos ahora la distribución límite para tales funciones aleatorias. Esto se hace cuando se cumpla la condición (R) del § 16. No obstante, para simplificar los razonamientos, introduzcamos ciertas

*) De la observación 25.2 resultará que (10) también es válida para u_n que crecen aún más lentamente.

suposiciones adicionales que no siempre están relacionadas con la esencia de la cuestión, pero hacen más breves y más claras las demostraciones.

Designemos con el símbolo (RR), las condiciones introducidas para indicar asimismo que tales son las condiciones de regularidad y que ellas intensifican las condiciones (R).

Condiciones (RR):

1) se *cumplen las condiciones* (A_0), (A_c), (R).

2) *la función* $l(x, \theta)$ *para c.t.* $[\mu]$ *valores de* x *es dos veces continuamente derivable respecto a* θ. *La función* $|l''(x, t)|$ *es mayorada por la función* $l(x)$ *que no depende de* t: $|l''(x, t)| < l(x)$, *para la cual la integral*

$$\mathbf{M}_t l(x_1) = \int l(x) f_t(x) \mu(dx)$$

converge uniformemente en $t \in \Theta$*).

Por convergencia uniforme de la integral entendemos la convergencia**)

$$\sup_\theta \int_{x:|l(x)|>N} l(x) f_\theta(x) \mu(dx) \to 0$$

cuando $N \to \infty$.

Posteriormente necesitaremos las dos propiedades siguientes, que se deducen de (RR):

1) *Validez de la derivación doble respecto al parámetro bajo el signo de integral en la igualdad*

$$\int f_\theta(x) \mu(dx) = 1$$

que significa la validez de las relaciones

$$\int f_0'(x) \mu(dx) = 0, \quad \int f_\theta''(x) \mu(dx) = 0. \tag{1}$$

2) *Convergencia uniforme de la integral*

$$I(\theta) = \int (l'(x, \theta))^2 f_\theta(x) \mu(dx).$$

(esta propiedad se deduce de (R) y se necesitará en el § 29).

---

*) Toda la exposición ulterior conservará su validez si la condición y la existencia de la mayorante se debilitan del modo siguiente: la región Θ puede ser cubierta por el número finito de regiones $\Theta_1, \ldots, \Theta_s$ de tal modo que cuando $\theta \in \Theta_j$ la función $l''(x, \theta)$ es mayorada por la función $l_{(j)}(x)$ que no depende de t: $|l''(x, \theta)| < l_{(j)}(x)$, para la cual la integral

$$\mathbf{M}_\theta l_{(j)}(x_1) = \int k_{(j)}(x) f_\theta(x) \mu(dx)$$

converge uniformemente en $\theta \in \Theta_j$, $j = 1, \ldots, s$.

**) Tal comprensión de la convergencia uniforme se halla en concordancia con la convergencia uniforme utilizada en el teorema 1.5.4. Aquí ella pertenecía a la función $l(x) = x$. A su vez, la misma no es la convergencia uniforme $\int \varphi(x, \theta) \mu(dx)$ para $\varphi(x, \theta) = l(x) f_\theta(x)$ cuando se supone que, para $N \to \infty$,

$$\sup_\theta \int_{x:|\varphi(x,\theta)|>N} (x, \theta) \mu(dx) \to 0.$$

Para descargar la exposición fundamental, la demostración de estos corolarios de las condiciones (RR) se da en el Suplemento VI. La exposición también se puede simplificar de otra manera: introduciendo en las condiciones (RR) las dos propiedades mencionadas y despreciando el hecho de que en tal forma ellas serán "redundantes".

En vista de que

$$l'(x,\ \theta) = \frac{f'_\theta(x)}{f_\theta(x)},\quad l''(x,\ \theta) = \frac{f''_\theta(x)}{f_\theta(x)} - \left(\frac{f'_\theta(x)}{f_\theta(x)}\right)^2,$$

la relación (1) se puede escribir en la forma

$$\mathbf{M}_\theta l'(\mathrm{x}_1,\ \theta) = 0,\ \mathbf{M}_\theta l''(\mathrm{x}_1,\ \theta) = -\mathbf{M}_\theta(l'(\mathrm{x}_1,\ \theta))^2 = -I(\theta). \tag{2}$$

Ya hemos utilizado la primera de estas igualdades.

Señalemos un corolario más de las condiciones (RR). Estas últimas son mucho más fuertes que las condiciones utilizadas en los §§ 21 y 23 y, por consiguiente, *tienen lugar todas las afirmaciones de los teoremas del* § 23 *acerca de las estimaciones para la distribución* $\sup\limits_{|v|\geq u} Z(v/\sqrt{n})$, *y acerca de la conciliabilidad de la e.v.m.*

Lema 1. *Si se cumplen las condiciones* (RR), *tiene lugar la continuidad* $l''(x,\ \theta)$ *"por término medio" desde el punto de vista siguiente*:

$$\mathbf{M}_\theta\omega''_\Delta(\mathrm{x}_1) = \int\omega''_\Delta(x)f_\theta(x)\mu(dx) \to 0 \tag{3}$$

para $\Delta\to 0$, *donde* $\omega''_\Delta(x)$ *es el módulo de continuidad de la función* $l''(x,\ \theta)$:

$$\omega''_\Delta(x) = \sup_{\substack{\theta\in\Theta,\theta+u\in\Theta \\ \in\Theta \\ |u|\leq\Delta}} |l''(x,\ \theta+u) - l''(x,\ \theta)|. \tag{4}$$

Demostración. En virtud del teorema de convergencia mayorable, la relación (3) será el corolario de la continuidad ordinaria, puesto que en este caso $\omega''_\Delta(x)\to 0$ para c.t. $[\mu]$ valores de x cuando $\Delta\to 0$ y, además, $|\omega''_\Delta(x)|\leqslant 2l(x)$. ◁

Designemos

$$\gamma_n(\Delta,\ \theta) = \sup_{|v|\leq\Delta}\left|\frac{L'(X,\ \theta+v) - L'(X,\ \theta)}{nv} + I(\theta)\right|.$$

Lema 2. *Supongamos que se cumplen las condiciones* (RR), $\delta_n > 0$, $n = 1, 2, \ldots$, *es cualquier sucesión convergente a cero. Entonces, para cualquier* $\theta\in\Theta$ *y para* $X\in\mathbf{P}_\theta$,

$$\gamma_n(\delta_n,\ \theta)\underset{\text{c.s.}}{\to} 0,\ \gamma_n(\delta_n,\ \hat\theta^*)\underset{\text{c.s.}}{\to} 0.$$

En estas relaciones, $I(\theta)$ *se puede sustituir por* $I(\hat\theta^*)$ *y al contrario.*

Demostración. Demostremos al principio la primera afirmación. Como $\mathbf{M}_\theta l''(x_1, \theta) = -I(\theta)$, $L''(X, \theta)/n \underset{c.s.}{\to} -I(\theta)$, es suficiente cerciorarse de que $\gamma_n(\delta_n) \to 0$, donde

$$\gamma_n(\Delta) = \sup_{|v| \leqslant \Delta} \left| \frac{L'(X, \theta + v) - L'(X, \theta)}{nv} - \frac{L''(X, \theta)}{n} \right|.$$

Pero

$$\gamma_n(\delta_n) = \leqslant \sup_{|v| \leqslant \delta_n} \frac{1}{n} |L''(X, \theta + v) - L''(X, \theta)| \leqslant \frac{1}{n} \sum_{i=1}^{n} \omega''_{\delta_n}(x_i) \equiv \bar{\omega}''_{\delta_n}(X),$$

donde $\omega''_\Delta(x)$ significa el módulo de continuidad $l''(x, \theta)$, definido en (4). Es evidente que para cualquier $\Delta > 0$ registrado, cuando n son bastante grandes,

$$\bar{\omega}''_{\delta_n}(X) \leqslant \bar{\omega}''_\Delta(X).$$

Además, según la ley fuerte de los grandes números,

$$\bar{\omega}''_\Delta(X) \underset{c.s.}{\to} \mathbf{M}_\theta \omega''_\Delta(x_1) \equiv \omega''_\Delta.$$

En virtud del lema 1, $\omega''_\Delta \to 0$ cuando $\Delta \to 0$. De aquí se deduce que

$$\bar{\omega}''_{\delta_n}(X) \underset{c.s.}{\to} 0. \qquad (5)$$

La primera afirmación queda demostrada. De (5) y de la definición de la convergencia casi segura se desprende que a la par con (5),

$$\bar{\omega}''_{\delta_n + \eta_n}(X) \underset{c.s.}{\to} 0$$

para toda sucesión de las variables aleatorias $\eta_n \underset{c.s.}{\to} 0$. Nos queda señalar que

$$\sup_{|v| < \delta_n} \left| \frac{L'(X, \hat{\theta}^* + v) - L'(X, \hat{\theta}^*)}{nv} - \frac{L''(X, \theta)}{n} \right| \leqslant \bar{\omega}''_{\delta_n + |\theta^* - \theta|}(X) \qquad (6)$$

y hacer uso del corolario 23.1. La posibilidad de sustituir $I(\theta)$ por $I(\hat{\theta}^*)$ también se deduce del corolario 23.1 (y de la continuidad de $I(\theta)$). ◁

Ahora podemos enunciar las principales afirmaciones acerca del comportamiento asintótico de la relación de verosimilitud $Z(t)$. Designemos

$$Y(u) = \ln Z(u/\sqrt{n}) = L(X, \theta + u/\sqrt{n}) - L(X, \theta)$$

y convengamos en designar por $\varepsilon_n(X, \theta)$ (a veces con índices adicionales) las diferentes sucesiones de variables aleatorias convergentes casi seguramente a cero respecto a $\mathbf{P}_\theta$.

Teorema 1. *Supongamos que se cumplen las condiciones (RR), $\delta_n > 0$ es una sucesión arbitraria que converge hacia el cero. Entonces para $|u/\sqrt{n}| < \delta_n$*

$$Y(u) = u\xi_n - \frac{u^2}{2} I(\theta)(1 + \varepsilon_n(X, \theta, u)), \tag{7}$$

donde

$$|\varepsilon_n(X, \theta, u)| \leqslant \varepsilon_n(X, \theta) \underset{\text{c.s.}}{\to} 0, \ \xi_n = L'(X, \theta)/\sqrt{n} \in \Phi_{0, I(\theta)}.$$

El punto $u^* = (\hat{\theta}^* - \theta)\sqrt{n}$, en el que $Y(u)$ alcanza el valor máximo, posee la propiedad

$$u^* = \frac{\xi_n}{I(\theta)} (1 + \varepsilon_n(X, \theta)), \tag{8}$$

$$2Y(u^*) = 2 \ln Z(\hat{\theta}^* - \theta) = \frac{\xi_n^2}{I(\theta)} (1 + \varepsilon_n(X, \theta)) \in H_1. \tag{9}$$

A la par con (7) es válida la representación

$$Y(u) = Y(u^*) - \frac{(u - u^*)^2}{2} I(\theta)(1 + \varepsilon_n(X, \theta, u)), \tag{10}$$

$$|\varepsilon_n(X, \theta, u)| < \varepsilon_n(X, \theta).$$

En todas las afirmaciones dadas se puede sustituir $I(\theta)$ por $I(\hat{\theta}^*)$.

En este teorema, al igual que en el lema 2, se supone que $\theta + u\sqrt{n} \in \Theta$. Esta relación será cumplida automáticamente para n bastante grandes si θ es el punto interior de Θ.

Observación. 1. Es importante notar que en (7) las variables aleatorias ξ_n y $\varepsilon_n(X, \theta)$ no dependen de n. Por eso la primera afirmación del teorema puede ser escrita en la forma

$$\sup_{|u| \leqslant \delta_n\sqrt{n}} \left| \frac{Y(u) - u\xi_n + \frac{u^2}{2} I(\theta)}{u^2} \right| \underset{\text{c.s.}}{\to} 0.$$

Si δ_n es tal que

$$\sum e^{-ng\delta_n^2/4} < \infty, \tag{11}$$

del teorema 23.2 se deduce que en la región adicional $|u| > \delta_n\sqrt{n}$,

$$\sup_{|u| > \delta_n\sqrt{n}} Y(u) \underset{\text{c.s.}}{\to} -\infty.$$

Demostración del teorema 1. Del lema 2 $|v| \leqslant \delta_n$ obtenemos

$$L'(X, \theta + v) = L'(X, \theta) - nvI(\theta)(1 + \varepsilon_n(X, \theta, v)),$$

$$|\varepsilon_n(X, \theta, v)| \leqslant \varepsilon_n(X, \theta).$$

Integrando esta igualdad respecto a v dentro de los límites de 0 a $u/\sqrt{n}$, obtendremos

$$L(X, \theta + u/\sqrt{n}) - L(X, \theta) = uL'(X, \theta)/\sqrt{n} - \frac{u^2}{2} I(\theta)(1 + \varepsilon_n(X, \theta, u)),$$

$$|\varepsilon_n(X, \theta, u)| \leqslant \varepsilon_n(X, \theta). \qquad (12)$$

Esto es, evidentemente, el desarrollo en serie de Taylor, donde $L''(X, \theta)/n$ ha sido sustituida por $I(\theta)$, y el término residual admite una estimación uniforme. En vista de que

$$\xi_n \equiv \frac{1}{\sqrt{n}} L'(X, \theta) = \frac{1}{\sqrt{n}} \sum l'(x_i, \theta)$$

es la suma de las variables independientes igualmente distribuidas, que tienen por media 0 y por varianza $I(\theta)$ (véase (2)), según el teorema central del límite $\xi \Subset \Phi_{0, I(\theta)}$. La representación (7) queda demostrada. Para demostrar (8) volvamos al lema (2). Este significa que existe un conjunto A, $\mathbf{P}_\theta(A) = 1$ tal que para $X_\infty \in A$, $n \to \infty$,

$$\sup_{|v| < \delta_n} \left| \frac{L'(X, \theta + v) - L'(X, \theta)}{nv} + I(\theta) \right| \to 0. \qquad (13)$$

Además, en virtud del corolario 23.1 existen la sucesión $u_n \to \infty$, $u_n/\sqrt{n} \equiv \gamma_n \to 0$ (u_n debe satisfacer (23.9)) y el conjunto B, $\mathbf{P}_\theta(B) = 1$ tal que para $X_\infty \in B$, $n \to \infty$,

$$v^* = (\hat{\theta}^* - \theta) = o(\gamma_n). \qquad (14)$$

Como la sucesión $\delta_n \to 0$ en (13) es arbitraria, para $X_\infty \in A \cap B$, $\mathbf{P}_\theta(A \cap B) = 1$, en virtud de (14) la relación (13) resultará justa en el punto $v = v^*$. Recordando que $L'(X, \theta + v^*) = L'(X, \hat{\theta}^*) = 0$, obtenemos para $X_\infty \in A \cap B$,

$$\left| I(\theta) - \frac{L'(X, \theta)}{n(\hat{\theta}^* - \theta)} \right| \to 0.$$

Esto significa que $\xi_n - I(\theta)u^* = u^*\varepsilon_n(X, \theta)$, y demuestra (8).

Haciendo uso de los mismos argumentos, se puede sustituir $u = u^* = v^*\sqrt{n} = (\hat{\theta}^* - \theta)\sqrt{n} = \frac{\xi_n}{I(\theta)} (1 + \varepsilon_n(X, \theta))$ en (12). Esto da

$$L(X, \hat{\theta}^*) - L(X, \theta) = \frac{\xi_n^2}{I(\theta)} (1 + \varepsilon_n(X, \theta))$$

y demuestra la primera parte de la relación (9). La convergencia de $\xi_n^2/I(\theta)$ hacia la distribución χ^2 con un grado de libertad se deduce de los teoremas de continuidad, ya que $\xi_n/\sqrt{I(\theta)} \Subset \Phi_{0,1}$.

La relación (10) se demuestra de un modo completamente análogo a (7) si se hace uso de la segunda afirmación del lema 2 y, basándose en ésta, se halla la representación para $L(X, \theta + u\sqrt{n}) - L(X, \hat{\theta}^*)$. ◁

Observación 2. En el lenguaje de las distribuciones, la primera afirmación del teorema 1 puede ser enunciada de la manera siguiente:

$$Y(u) \in \Phi_{-u^2 I(\theta)/2,\ u^2 I(\theta)}. \tag{15}$$

Anteriormente hemos señalado que la segunda condición (RR) (acerca de la existencia de $l''(x, \theta)$) no siempre es esencial para las afirmaciones que han de ser demostradas. El carácter no esencial de esta condición para la convergencia (15) se puede mostrar mediante los razonamientos siguientes. La magnitud

$$Y(u) = L\left(X, \theta + \frac{u}{\sqrt{n}}\right) - L(X, \theta) = \sum_{i=1}^{n}\left[l\left(x_i, \theta + \frac{u}{\sqrt{n}}\right) - l(x_i, \theta)\right]$$

es la suma de las magnitudes independientes igualmente distribuidas. Por eso, según el teorema central del límite para el esquema de series (los sumandos dependen de n y omitimos la verificación de las condiciones de Lindeberg)

$$Y(u) \in \Phi_{\alpha(u),\ \sigma^2(u)},$$

donde

$$\alpha(u) = \lim_{n\to\infty} n\mathbf{M}_\theta[l(x_1, \theta + u/\sqrt{n}) - l(x_1, \theta)] =$$

$$= \lim_{n\to\infty} n\mathbf{M}_\theta \ln \frac{f_{\theta+u/\sqrt{n}}(x_1)}{f_\theta(x_1)} = -u^2 \lim_{\Delta\to 0} \frac{\varrho_1(\mathbf{P}_{\theta+\Delta}, \mathbf{P}_\theta)}{\Delta^2} = -u^2 I(\theta)/2$$

(véase el teorema 21.2 y la observación 21.1). Luego

$$\sigma^2(u) = \lim_{n\to\infty} n\mathbf{M}_\theta[l(x_1, \theta + u/\sqrt{n}) - l(x_1, \theta)]^2 =$$

$$= u^2 \lim_{\Delta\to 0} \int \left[\frac{l(x, \theta + \Delta) - l(x, \theta)}{\Delta}\right]^2 f_\theta(x)\mu(dx) =$$

$$= u^2 \int (l'(x, \theta))^2 f_\theta(x)\mu(dx) = u^2 I(\theta).$$

Si al calcular $\alpha(u)$ y $\sigma^2(u)$ se utilizó el desarrollo $l(x, \theta + u/\sqrt{n})$ en serie con dos derivadas, obtendríamos el mismo resultado. Sin embargo, nos hemos cerciorado de que no es obligatorio hacer esto.

Concluyendo este párrafo, del teorema 1 obtendremos otro corolario útil que necesitaremos en adelante y que se refiere al comportamiento de las integrales de la relación de verosimilitud.

Teorema 2. *Supongamos que se cumplen las condiciones* (RR), *la función* $w(t)$ *satisface la condición*

$$|w(t)| \leqslant ce^{\alpha|t|^2}, \quad c < \infty, \quad \alpha = g/16$$

($g > 0$ está definido en el § 21) *y la función $q(t)$ es continua en el punto $t = 0$ y está limitada. Supongamos, además, que Π es cualquier medida en $(R, \mathfrak{B})$, tal que $\int e^{-\alpha|u|^2/4}\Pi(du) < \infty$. En este caso, si θ es un punto interior de Θ* y $X \in \mathbf{P}_\theta$,

$$J \equiv \int w(u^* - u)q(\theta + u/\sqrt{n})Z(u/\sqrt{n})\Pi(du) =$$

$$= e^{Y(u^*)}q(\theta)\left[\int w(u^* - u)e^{-\frac{1}{2}(u-u^*)^2 I(\theta)}\Pi(du) + \varepsilon_n(X, \theta)\right]. \qquad (16)$$

En particular, si Π es la medida de Lebesgue, $\Pi(du) = du$, entonces

$$J = \sqrt{\frac{2\pi}{I(\theta)}}\, e^{Y(u^*)}q(\theta)(\mathbf{M}w(\eta) + \varepsilon_n(X, \theta)),$$

$$donde\ \varepsilon_n(X, \theta) \underset{c.s.}{\rightarrow} 0,\ \eta \in \Phi_{0,\, I^{-1}(\theta)}.$$

La afirmación (16) es muy natural, ya que el factor $q(\theta + u/\sqrt{n})$ es "casi constante" y la función $Z(u/\sqrt{n}) = e^{Y(u)}$ se aproxima, con una exactitud de hasta el factor constante, según el teorema 1, con una densidad de distribución normal.

Demostración. Para simplificar la notación nos limitaremos a examinar el caso cuando Π es la medida de Lebesgue. El paso al caso general no presenta ninguna dificultad.

Estimemos primeramente la parte de la integral (16) en la región $|u| > r$. Designémosla por $J(r)$. Como $f_\theta(X)/f_{\theta^*}(X) \leqslant 1$, entonces, suponiendo, para abreviar, $Z = Z(u^*/\sqrt{n}) = e^{Y(u^*)}$, $t = \theta + u/\sqrt{n}$, obtenemos

$$Z^{-1}Z\left(\frac{u}{\sqrt{n}}\right) = \frac{f_t(X)}{f_{\theta^*}(X)} \leqslant \left(\frac{f_t(X)}{f_{\theta^*}(X)}\right)^{3/4} \leqslant Z^{3/4}\left(\frac{u}{\sqrt{n}}\right).$$

Por eso, en virtud de la desigualdad de Cauchy — Buniakovski, del teorema 23.1 y del corolario 23.1,

$$\mathbf{M}_\theta w(u^* - u)Z^{-1}Z(u/\sqrt{n}) \leqslant$$

$$\leqslant [\mathbf{M}_t w^2(\sqrt{n}(\theta^* - t))\mathbf{M}_\theta Z^{1/2}(u/\sqrt{n})]^{1/2} \leqslant ce^{-u^2 g/4}.$$

Como máx $q(t) < \infty$, de aquí y del lema 23.1 hallamos

$$\mathbf{M}_\theta Z^{-1}J(r) \leqslant ce^{-gr^2/4}.$$

Haciendo uso de la desigualdad de Chébishev, obtenemos las estimaciones del mismo orden también para $\mathbf{P}_\theta(Z^{-1}J(r) > \delta)$. Por eso, si $r = r_n \to \infty$, de modo que

$$\sum e^{-r_n^2 g/4} < \infty, \qquad (17)$$

entonces, para $y \geqslant r_n$,

$$Z^{-1}J(y) \underset{c.s.}{\to} 0. \tag{18}$$

Elijamos $r_n = o(\sqrt{n})$ y examinemos la parte restante de la integral $V(y) = J - J(y)$ cuando $y = 2r_n$. Según el teorema 1,

$$Z^{-1}V(2r) = Z^{-1} \int\limits_{|u|<2r_n} q(\theta + u/\sqrt{n})w(u^* - u)Z(u/\sqrt{n})du =$$

$$= \int\limits_{|u|<2r_n} (q(\theta) + \varepsilon_n(u))w(u^* - u) \times$$

$$\times \exp\left\{-\frac{1}{2}(u - u^*)^2 I(\theta)(1 + \varepsilon_n(X, \theta, u))\right\}du,$$

donde $|\varepsilon_n(u)| < \varepsilon_n \to 0$, $|\varepsilon_n(X, \theta, u)| \leqslant \varepsilon_n(X, \theta) \underset{c.s.}{\to} 0$ cuando $n \to \infty$. Por eso, en virtud de (18), es suficiente cerciorarse de la proximidad de las integrales

$$\int\limits_{|u|<2r_n} w(u^* - u) \exp\left\{-\frac{1}{2}(u - u^*)^2 I(\theta)(1 + \varepsilon_n(X, \theta, u))\right\}du,$$

$$\sqrt{\frac{2\pi}{I(\theta)}}\, \mathbf{M}w(\eta) = \int w(u^* - u) \exp\left\{-\frac{1}{2}(u - u^*)^2 I(\theta)\right\}du.$$

En virtud de (17) y del corolario 23.1 existe un conjunto A, $\mathbf{P}_\theta(A) = 1$ tal, que $|u^*| \leqslant r_n$ para $X_\infty \in A$ cuando todos $n = n(X_\infty)$ son bastante grandes. Como $I(\theta) \geqslant g$, $|u - u^*|^2 > u^2/2$ para $|u| > 2r_n$, $|u^*| < r_n$, entonces, en el conjunto A (véase el lema 23.1),

$$\int\limits_{|u|\geqslant 2r_n} w(u^* - u) \exp\left\{-\frac{1}{2}(u - u^*)^2 I(\theta)\right\}du < ce^{-gr_n^2} \to 0.$$

Por eso nos queda estimar

$$\int\limits_{|u|<2r_n} w(u^* - u)\left|\exp\left\{-\frac{1}{2}(u - u^*)^2 I(\theta)(1 + \varepsilon_n(X, \theta, u))\right\} - \right.$$

$$\left. - \exp\left\{-\frac{1}{2}(u - u^*)^2 I(\theta)\right\}\right|du \leqslant \int w(v)\left|\exp\left\{-\frac{1}{2}v^2 I(\theta) \times \right.\right.$$

$$\left.\left. - (1 + \varepsilon_n(X, \theta, v + u^*))\right\} - \exp\left\{-\frac{1}{2}v^2 I(\theta)\right\}\right|dv.$$

Pero esta integral converge en el conjunto AB hacia el cero, donde $B = \{X_\infty: \varepsilon_n(X, \theta) \to 0\}$, $\mathbf{P}_\theta(B) = 1$. Esto resulta de la convergencia a cero para cada v de la función subintegral y del hecho de que ésta es mayorada por la función sometida a integración. $\triangleleft$

§ 25. Propiedades de las estimaciones de verosimilitud máxima. Normalidad asintótica. Optimación asintótica

Supongamos que $X \in \mathbf{P}_\theta$ y $\hat{\theta}^*$ es la e.v.m. Los resultados de los párrafos precedentes permiten describir por completo las propiedades asintóticas de θ^* cuando el volumen n de la muestra crece indefinidamente. Además, en este párrafo hemos establecido uno de los resultados centrales del capítulo presente, que consiste en que la e.v.m, al cumplirse las condiciones (RR), posee todas las propiedades posibles de optimación asintótica, que hemos examinado anteriormente, o sea, la estimación asintóticamente eficiente es, a la vez, asintóticamente bayesiana (para toda distribución a priori que tiene densidad) y asintóticamente minimax.

En este párrafo siempre supondremos, sin especificarlo complementariamente, que se cumplen las condiciones (RR).

1. Normalidad asintótica de la e.v.m.

Teorema 1. *La e.v.m. $\hat{\theta}^*$ es una estimación asintóticamente normal, con la particularidad de que la convergencia*

$$u^* = (\hat{\theta}^* - \theta)\sqrt{n} \Subset \Phi_{0,\, I^{-1}(\theta)} \tag{1}$$

tiene lugar junto con los momentos de cualquier orden, o sea, junto con (1), *para cualquier $k > 0$, se cumple*

$$\mathbf{M}_\theta (u^*)^k \to \mathbf{M}\eta^k, \quad \eta \in \Phi_{0,\, I^{-1}(\theta)}. \tag{2}$$

Además, para cualquier función continua $w(t)$ tal, que $|w(t)| < e^{gt^2/6}$ (véase (23.4)),

$$\mathbf{M}_\theta w(u^*) \to \mathbf{M} w(\eta), \quad \eta \in \Phi_{0,\, I^{-1}(\theta)}. \tag{3}$$

Demostración. En el teorema 24.1 hemos establecido que

$$u^* = (\hat{\theta}^* - \theta)\sqrt{n} = \frac{\xi_n}{I(\theta)}(1 + \varepsilon_n(X, \theta)), \tag{4}$$

donde $\varepsilon_n(X, \theta) \underset{c.s.}{\to} 0$, $\xi_n = L'(X, \theta)/\sqrt{n} \Subset \Phi_{0, I(\theta)}$. Esto demuestra (1). Las relaciones (2) y (3) se obtienen de (1) y del teorema de continuidad para los momentos (véase el § 1.5), puesto que en virtud del corolario 23.2,

$$\mathbf{M}_\theta w^{6/5}(u^*) \leqslant \mathbf{M}_\theta \exp\left\{\frac{(u^*)^2 g}{5}\right\} < c < \infty. \quad \triangleleft$$

Observación 1. De (1) y (2) se deduce que $\hat{\theta}^*$ pertenece a la clase de estimaciones $\mathbf{K}_{\Phi,2}$, en la que la convergencia de $(\hat{\theta}^* - \theta)\sqrt{n} \Subset \Phi_{0,\sigma^2(\theta)}$ tiene lugar junto con la convergencia de $\mathbf{M}_\theta(\theta^* - \theta)^2 \to \sigma^2(\theta)$ de los primeros momentos. Como ya hemos señalado en el § 8, en esta clase, el enfoque asintótico de la comparación de las estimaciones coincide, de hecho, con el enfoque estándar.

Observación 2. La relación (4) también permite describir exactamente las "desviaciones máximas" de $(\hat{\theta}^* - \theta)\sqrt{n}$ cuando $n \to \infty$. Pues, se sabe (véanse [61] y [84]) que las sumas normalizadas ξ_n de las magnitudes independientes igualmente distribuidas, que tienen por media el cero y por varianza $I(\theta)$, satisfacen la ley de logaritmo reiterado, en virtud de la cual

$$\mathbf{P}\left(\lim_{n\to\infty}\sup \frac{|\xi_n|}{\sqrt{2I(\theta)\ln\ln n}} = 1\right) = 1.$$

En vista de que en (4) $\lim\sup_{n\to\infty} \varepsilon_n(X, \theta) = 0$ c.s., obtenemos que

$$\mathbf{P}_\theta\left(\lim_{n\to\infty}\sup \frac{|\theta^* - \theta|\sqrt{nI(\theta)}}{\sqrt{2\ln\ln N}} = 1\right) = 1.$$

Determinemos ahora, en calidad de corolarios del teorema 2, algunas propiedades de la e.v.m. relacionadas con la optimación asintótica.

2. Eficacia asintótica. En el § 16 hemos introducido el estudio de la clase $\tilde{K}_0$ de estimaciones asintóticamente no desplazadas, o sea, de estimaciones θ^* cuyo desplazamiento $b(\theta) = \mathbf{M}_\theta\theta^* - \theta$ posee las propiedades

$$b(\theta) = o(1/\sqrt{n}, \; b'(\theta) = o(1). \tag{5}$$

En el § 20 hemos expuesto las ideas según las cuales, en búsqueda de las estimaciones asintóticamente eficientes "en total", es posible limitarse a la clase $\tilde{K}_0$.

Establezcamos ahora el hecho siguiente.

Corolario 1. $\theta^* \in \tilde{K}_0$.

Demostración. La primera de las relaciones (5) resulta de (2) cuando $k = 1$. Para demostrar la segunda señalemos que (véase el § 16)

$$1 + b'(\theta) = \mathbf{M}_\theta\hat{\theta}^* L'(X, \theta) = \mathbf{M}_\theta(\hat{\theta}^* - \theta)L'(X, \theta) =$$

$$= \mathbf{M}_\theta((\hat{\theta}^* - \theta)\sqrt{n}\,\xi_n) = \mathbf{M}_\theta \frac{\xi_n^2}{I(\theta)}(1 + \varepsilon_n(X, \theta)),$$

$$\varepsilon_n(X, \theta) \underset{\text{c.s.}}{\to} 0.$$

Si aquí es cierto el teorema de continuidad para los momentos, entonces obtenemos la relación requerida $1 + b'(\theta) \to 1$ o, que es lo mismo, $b'(\theta) \to 0$. Para establecer la validez de este teorema en nuestro caso, es

suficiente cerciorarse (véase el § 1.5) de que

$$\mathbf{M}_\theta|(\hat{\theta}^* - \theta)\sqrt{n}\,\xi_n|^{3/2} < c < \infty, \tag{6}$$

donde c no depende de n. Hagamos uso de la desigualdad de Hölder

$$\mathbf{M}|\xi\eta|^r \leqslant (\mathbf{M}|\xi|^{pr})^{1/p}(\mathbf{M}|\eta|^{qr})^{1/q}, \quad p > 0, \quad q > 0, \quad \frac{1}{p} + \frac{1}{q} = 1$$

para $r = 3/2$, $p = 4$, $q = 4/3$. Entonces obtenemos, para el primer miembro de (6), la estimación $(\mathbf{M}_\theta[(\hat{\theta}^* - \theta)\sqrt{n}]^6)^{1/4}(\mathbf{M}\xi_n^2)^{3/4}$, que, en virtud de (2), nos da la desigualdad deseada. ◁

El corolario siguiente, debido a su importancia, lo enunciaremos en forma de teorema.

Teorema 2. *La e.v.m. $\hat{\theta}^*$ es una estimación asintóticamente R-eficiente. Además, $\hat{\theta}^*$ es asintóticamente eficiente en $\mathfrak{K}_0$.*

Demostración. El hecho de que $\hat{\theta}^*$ es una estimación asintóticamente R-eficiente se desprende directamente de la definición 16.1 y del hecho de que

$$\mathbf{M}_\theta(\hat{\theta}^* - \theta)^2 = \frac{1 + o(1)}{nI(\theta)}.$$

La eficacia asintótica en $\mathfrak{K}_0$ se deduce del teorema 16.3. ◁

El teorema 2, junto con las observaciones referentes al teorema 16.3, significa que, al cumplirse las condiciones (RR), cualquier estimación asintóticamente eficiente en $\mathfrak{K}_0$ será una estimación asintóticamente R-eficiente.

Anotemos que la contracción del conjunto de las estimaciones examinadas, hasta $\mathfrak{K}_0$, no es la única contracción, ni mucho menos, con la que $\hat{\theta}^*$ se vuelve asintóticamente eficiente.

Indiquemos otra contracción relacionada en este caso con la propiedad de θ de ser mediana asintótica de la distribución de las estimaciones asintóticamente normales, o sea, con la propiedad

$$\mathbf{P}_\theta(\hat{\theta}^* > \theta) \to 1/2 \tag{7}$$

cuando $n \to \infty$.

Designemos por $\mathfrak{K}^\circ$ la clase de estimaciones θ^* para las cuales (7) se cumple uniformemente respecto a θ. La clase $\mathfrak{K}^\circ$ podría llamarse clase de *estimaciones asintóticamente centrales*.

Teorema 3. *La e.v.m. $\hat{\theta}^* \in \mathfrak{K}^\circ$ es precisamente una estimación asintóticamente eficiente en la clase $\mathfrak{K}^\circ$.*

Aplazaremos la demostración de este teorema hasta el § 3.3.

3. Carácter asintóticamente bayesiano de la e.v.m. En este apartado, por doquier se suponga la existencia de la densidad $q(t)$ de la distribución a

priori Q respecto a la medida de Lebesgue en Θ, supondremos también, sin especificarlo complementariamente, que la densidad es integrable según Riemann, así que se satisfarán las condiciones del teorema 20.5.

Teorema 4. *La e.v.m. $\hat{\theta}^*$ es una estimación asintóticamente R-bayesiana. Si Q es una distribución arbitraria a priori que tiene una densidad $q(t)$ respecto a la medida de Lebesgue, entonces $\hat{\theta}^*$ también es una estimación asintóticamente bayesiana que* corresponde a la distribución Q.

Demostración. El carácter asintóticamente R-bayesiano de la e.v.m. se deduce de las relaciones

$$\lim_{n\to\infty} \mathbf{M}[\sqrt{n}(\hat{\theta}^* - \theta)]^2 = \lim_{n\to\infty} \mathbf{M}\mathbf{M}_\theta[\sqrt{n}(\hat{\theta}^* - \theta)]^2 =$$

$$= \mathbf{M} \lim_{n\to\infty} \mathbf{M}_\theta[\sqrt{n}(\hat{\theta}^* - \theta)]^2 = \mathbf{M}I^{-1}(\theta) = J.$$

Aquí el paso límite bajo el signo de la esperanza matemática es legítimo según el teorema de la convergencia mayorada, ya que, en virtud de 23.2, el valor de $\mathbf{M}_\theta[\sqrt{n}(\hat{\theta}^* - \theta)]^2$ está uniformemente limitado por la constante que no depende de n ni de θ.

El carácter asintóticamente bayesiano se deduce del corolario 20.1. ◁

De las observaciones referentes al corolario 20.1 y del teorema 4 resulta que cualquier estimación asintóticamente bayesiana es asintóticamente R-bayesiana.

La afirmación del teorema 4 puede ser amplificada. Resulta que la e.v.m. y la estimación bayesiana "casi" coinciden para cualquier densidad a priori q.

Teorema 5.

$$\mathbf{M}n(\hat{\theta}^* - \theta_Q^*)^2 \to 0, \qquad (\theta_Q^* - \hat{\theta}^*)\sqrt{n} \xrightarrow[p]{} 0,$$

donde θ_Q^* es la estimación bayesiana que corresponde a la distribución Q, y la convergencia en probabilidad se entiende respecto a la distribución compatible de X y θ en $\mathscr{X}^n \times \Theta$.

El teorema 5 se desprende directamente del corolario 20.2. Su afirmación es equivalente a que para casi todos t

$$\mathbf{M}_t n(\hat{\theta}^* - \theta_Q^*)^2 \to 0.$$

Es posible la amplificación ulterior de la afirmación enunciada.

Teorema 6. *Sea θ un punto interior arbitrario Θ, $X \in \mathbf{P}_\theta$. Sea, luego, $q(t)$ una densidad arbitraria, continua y positiva dentro de Θ, de la distribución a priori. Entonces $\sqrt{n}(\hat{\theta}^* - \theta_Q^*) \xrightarrow[c.s.]{} 0$.*

La demostración de deduce del teorema 2 del párrafo precendente. En efecto, $\theta_Q^* - \hat{\theta}^* = \dfrac{\int (t - \hat{\theta}^*)q(t)f_t(X)dt}{\int q(t)f_t(X)dt}$. Sustituyendo las variables $t = \theta + u/\sqrt{n}$ y dividiendo por $f_\theta(X)$ el numerador y denominador en esta expresión, obtenemos

$$\theta_Q^* - \hat{\theta}^* = \frac{\int (u - u^*)q(\theta + u/\sqrt{n})Z(u/\sqrt{n})du}{\sqrt{n}\int q(\theta + u/\sqrt{n})Z(u/\sqrt{n})du}.$$

Ahora es necesario hacer uso del teorema 24.2 para $w(t) = t$ y $w(t) = 1$. Como en el primer caso $\mathbf{M}w(\eta) = \mathbf{M}\eta = 0$, entonces obtenemos

$$\theta_Q^* - \hat{\theta}^* = \varepsilon_n(X, \theta)/\sqrt{n}, \quad \varepsilon_n(X, \theta) \underset{c.s.}{\to} 0. \ \triangleleft$$

4. Carácter asintóticamente minimax de la e.v.m.

Teorema 7. *La e.v.m. es una estimación asintóticamente minimax.*

Este teorema se deduce directamente del corolario 20.3 y de la afirmación siguiente.

Lema 1.

$$\lim_{n\to\infty} \sup_{\theta\in\Gamma} \mathbf{M}_\theta n(\hat{\theta}^* - \theta)^2 = \sup_{\theta\in\Gamma} I^{-1}(\theta),$$

donde Γ *es cualquier trazado dentro de* Θ.

El lema 1 se desprende de la convergencia (2) uniforme en θ. La uniformidad será demostrada en el § 29 (véase el apartado 29.3).

§ 26*. Cálculo aproximado de las estimaciones de verosimilitud máxima

Hemos visto que en los problemas de estimación de los parámetros revisten el máximo interés las estimaciones eficientes y asintóticamente eficientes y, en particular, las e.v.m. Surge la cuestión acerca de la determinación práctica de tales estimaciones. En los problemas reales, la búsqueda del valor exacto de la e.v.m. $\hat{\theta}^*$ puede presentar grandes dificultades. Esto se refiere, sobre todo, a las distribuciones que no tienen estadísticas suficientes relativamente sencillas.

Por otro lado, la determinación de cualquier estimación asintóticamente normal θ^* no provoca, por regla general, dificultades.

Aquí mostraremos un método de construcción de la estimación θ_1^*, asintóticamente equivalente a la e.v.m. $\hat{\theta}^*$ (y, por consiguiente, a la asintóticamente eficiente), el cual se basa en el método de Newton para cálculos

aproximados y en la utilización de la estimación asintóticamente normal θ^*. Pongamos

$$U(t) = t - L'(X, t)\cdot(L''(X, t))^{-1}, \quad t \in \Theta,$$
$$U_1(t) = t + L'(X, t)\cdot(nI(t))^{-1}, \quad t \in \Theta.$$

Teorema 1. *Supongamos que se cumplen las condiciones (RR), $X \in \mathbf{P}_\theta$ y que θ^* es cualquier estimación asintóticamente normal*

$$(\theta^* - \theta)\sqrt{n} \in \Phi_{0,\, \sigma^2(\theta)}.$$

En este caso la estimación $\theta_1^* = U(\theta^*)$ (o bien $\theta_1^* = U_1(\theta^*)$) será asintóticamente equivalente a $\hat{\theta}^*$, o sea,

$$(\theta_1^* - \hat{\theta}^*)\sqrt{n} \underset{P_\theta}{\to} 0.$$

La demostración del teorema se apoyará en el lema siguiente.

Lema 1. *Supongamos que se cumplen las condiciones (RR), $X \in \mathbf{P}_\theta$, y que $\delta_n > 0$ es una sucesión arbitraria convergente a cero. En este caso, si θ_n es tal que $|\theta_n - \theta| \leqslant \delta_n$,*

$$U(\theta_n) - \hat{\theta}^* = (\theta_n - \hat{\theta}^*)\varepsilon_n(\theta_n, \theta, X),$$

donde $\bar{\varepsilon}_n = \max\limits_{\theta_n: |\theta_n - \theta| \leqslant \delta_n} |\varepsilon_n(\theta_n, \theta, X)| \underset{P_\theta}{\to} 0.$

Esa misma afirmación será válida si en vez de U utilizamos la función U_1.

Con otras palabras, si se hace uso del método de aproximaciones sucesivas hacia $\hat{\theta}^*$ y se pone $\theta_0^* = \theta_n$, $\theta_1^* = U(\theta_0^*)$ (o bien $\theta_1^* = U_1(\theta_0^*)$), entonces $\theta_1^* - \hat{\theta}^* = o(\theta_0^* - \hat{\theta}^*)$, así que la aproximación θ_1^* es mucho mejor que θ_0^*.

Demostración. De las investigaciones de § 24 y de la continuidad de L'' se deduce (véase, por ejemplo, el lema 24.1) que

$$L'(X, \theta_n) = (\theta_n - \hat{\theta}^*)L''(X, \tilde{\theta}),\; L''(X, \tilde{\theta}) = n(I(\theta) + \varepsilon_n'(\theta_n, \theta, X)),$$

donde $\tilde{\theta} \in [\theta_n, \hat{\theta}^*]$, $\max\limits_{\theta_n: |\theta_n - \theta| \leqslant \delta_n} \varepsilon_n'(\theta_n, \theta, X) \underset{P_\theta}{\to} 0$ para cualquier sucesión $\delta_n \to$ $\to 0$. Luego,

$$L''(X, \theta_n) = n(I(\theta) + \varepsilon_n''),$$
$$(I(\theta) + \varepsilon_n')(I(\theta) + \varepsilon_n'')^{-1} = 1 + \varepsilon_n,$$

donde ε_n'', ε_n poseen la misma propiedad que ε_n'. Por consiguiente,

$$U(\theta_n) - \hat{\theta}^* = \theta_n - \theta^* - L'(X, \theta_n)(L''(X, \theta_n))^{-1} =$$
$$= \theta_n - \hat{\theta}^* - (\theta_n - \theta^*)(1 + \varepsilon_n) = (\theta_n - \hat{\theta}^*)\varepsilon_n.$$

La demostración para la función U_1 se realiza exactamente igual. ◁

Demostración del teorema 1. Elijamos cualquier $\delta_n \to 0$ tal, que $\delta_n\sqrt{n} \to \infty$, y representemos $(\theta_1^* - \hat{\theta}^*)\sqrt{n}$ en la forma

$$(U(\theta^*) - \hat{\theta}^*)\sqrt{n} = \sqrt{n}(\theta^* - \hat{\theta}^*)\varepsilon_n(\theta^*, \theta, X)I_{(|\theta^* - \theta| < \delta_n)} + r_n,$$

donde $r_n \neq 0$ únicamente en el conjunto $B_n = \{X: |\theta^* - \theta| > \delta_n\}$ y, en virtud del lema 1,

$$\bar{\varepsilon}_n = \max_{|t - \theta| \leq \delta_n} \varepsilon_n(t, \theta, X) \underset{P_\theta}{\to} 0.$$

Como, además, $\mathbf{P}_\theta(B_n) \to 0$, de aquí se deduce que

$$|\theta_1^* - \hat{\theta}^*|\sqrt{n} \leq \sqrt{n}|\theta^* - \theta|\bar{\varepsilon}_n + \sqrt{n}|\hat{\theta}^* - \theta|\bar{\varepsilon}_n + r_n \underset{P_\theta}{\to} 0. \quad \lhd$$

El teorema 1 muestra que el método de aproximaciones sucesivas, partiendo de cualquier estimación asintóticamente normal, nos lleva en 1 paso al punto θ^*, con una exactitud de hasta los valores de $o(1/\sqrt{n})$.

Si se exige la existencia de las terceras derivadas continuas $l'''(x, \theta)$, entonces también se puede comenzar de puntos más lejos, que distan de θ, digamos, a la magnitud de $o(n^{-1/4})$. En este caso, al igual que en las condiciones del teorema 1, en 1 paso resultaremos en el $o(1/\sqrt{n})$-entorno del punto $\hat{\theta}^*$. En efecto,

$$L'(X, t) = (t - \hat{\theta}^*)L''(X, \hat{\theta}^*) + \frac{(t - \hat{\theta}^*)^2}{2} L'''(X, \theta') =$$

$$= (t - \hat{\theta}^*)L''(X, t) + \frac{3}{2}(t - \hat{\theta}^*)^2 L'''(X, \theta''),$$

donde θ' y θ'' están comprendidos entre t y $\hat{\theta}^*$. Por eso

$$U(\theta_n) - \hat{\theta}^* = \theta_n - \hat{\theta}^* - L'(X, \theta_n)(L''(X, \theta_n))^{-1} =$$

$$= \frac{3}{2}(\theta_n - \hat{\theta}^*)^2(I(\theta) + \varepsilon_n), \quad \sqrt{n}(U(\theta_n) - \hat{\theta}^*) \underset{P_\theta}{\to} 0$$

si $|\theta_n - \theta| = o(n^{-1/4})$. $\lhd$

Ejemplo 1. *Clasificación de las partículas.* Examinemos una fuente que emite partículas de dos tipos: con probabilidad p, partículas del tipo A; y con probabilidad 1, p partículas del tipo B. La energía de las partículas es aleatoria y tiene una densidad de $f_1(x)$ para las partículas del tipo A, y de $f_2(x)$ para las del tipo B. Las funciones $f_i(x)$ son conocidas. Han sido registradas n partículas con energías $x_1, \ldots, x_n$. ¿A qué es igual la probabilidad p? Aquí la función de verosimilitud es igual a

$$f_p(X) = \prod_{i=1}^{n} (pf_1(x_i) + (1 - p)f_2(x_i)),$$

así que

$$L'(X, p) = \sum_{i=1}^{n} \frac{f_1(x_i) - f_2(x_i)}{pf_1(x_i) + (1 - p)f_2(x_i)}. \qquad (1)$$

Vemos que la búsqueda de la e.v.m. $\hat{p}^*$ conduce a la ecuación $L' = 0$ de grado $n - 1$ respecto a p, la cual se resuelve, para grandes n, con mucha dificultad. Hagamos uso del teorema 1. Para eso necesitamos cualquier estimación asintóticamente normal p^*. Supongamos que $\int (F_1 - F_2)^2 dx < \infty$, donde $F_i(x) = \int_{-\infty}^{x} f_i(t)dt$, y examinemos el enfoque natural siguiente. Definamos p^* como valor que minimiza

$$\int (F_n^*(x) - F(x))^2 dx, \quad F(x) = pF_1(x) + (1 - p)F_2(x). \tag{2}$$

Igualando a cero la derivada de (2), obtenemos $\int (F_n^* - F)(F_1 - F_2)dx = 0$,

$$p^* = \frac{\int (F_n^* - F_2)(F_1 - F_2)dx}{\int (F_1 - F_2)^2 dx}.$$

Es fácil notar que $\mathbf{M}p^* = p$ y que

$$(p^* - p)\sqrt{n} = \frac{\int (F_n^* - F)\sqrt{n}(F_1 - F_2)dx}{\int (F_1 - F_2)^2 dx}. \tag{3}$$

De los resultados de los §§ 1.6—1.8 se deduce que p^* es una estimación asintóticamente normal y que la distribución límite (3) coincide con la distribución

$$\frac{\int w^\circ(F(x))(F_1 - F_2)dx}{\int (F_1 - F_2)^2 dx}.$$

Por lo tanto, en virtud del teorema 1 la estimación

$$p_1^* = p^* - L'(X, p^*)(L''(X, p^*))^{-1},$$

donde L' está definida en (1),

$$L'' = -\sum \frac{(f_1(x_i) - f_2(x_i))^2}{(pf_1(x_i) + (1 - p)f_2(x_i))^2},$$

será asintóticamente equivalente a la e.v.m. $\hat{p}^*$. El coeficiente de dispersión p_1^* será determinado por la información

$$I(p) = \int \frac{(f_1(x) - f_2(x))^2}{pf_1(x) + (1 - p)f_2(x)} dx$$

y será menor que el coeficiente de dispersión p^*.

Ejemplo 2. Le proponemos al lector que halle, de ese mismo modo, la aproximación para la e.v.m. del parámetro α de la distribución de Cauchy

$K_{\alpha,1}$ que tiene una densidad de

$$k_{\alpha,1}(x) = \frac{1}{\pi(1 + (x - \alpha)^2)} .$$

En calidad de estimación asintóticamente normal "previa" se puede tomar la mediana muestral ζ^* (véase el § 2 ó los §§ 1.3 y 1.8 Aquí no se puede tomar la estimación $\alpha^* = \bar{x}$, ya que $M_\alpha\alpha^*$ no existe). La estimación

$$\alpha_1^* = \zeta^* - L'(X, \zeta^*)(L''(X, \zeta^*))^{-1},$$

donde
$$L'(X, \alpha) = -2 \sum \frac{x_i - \alpha}{1 + (x_i - \alpha)^2},$$

$$L''(X, \alpha) = 2 \sum \frac{1 - (x_i - \alpha)^2}{(1 + (x_i - \alpha)^2)^2},$$

será asintóticamente equivalente a la e.v.m. $\hat{\alpha}^*$. Como

$$I(\alpha) = \int \frac{(k'_{\alpha,1}(x))^2}{k_{\alpha,1}(x)}\,dx = \frac{4}{\pi}\int \frac{x^2}{(1 + x^2)^3}\,dx = \frac{1}{2},$$

los coeficientes de dispersión ζ^* y α_1^* serán iguales respectivamente (véase el § 2) a

$$\frac{1}{2k_{\alpha,1}(\alpha)} = \frac{\pi}{2}, \quad I^{-1/2}(\alpha) = \sqrt{2}, \quad \frac{\pi}{2} > \sqrt{2}.$$

Ejemplo 3. La sangre de cada persona pertenece a uno de los cuatro grupos que designamos por 0 (cero), A, B y AB. El heredamiento de los grupos de sangre es controlado por tres genes: A, B y 0, además, el gene 0 es "deprimido" por los genes A y B. Por eso, si p, q y $r = 1 - p - q$ designan las probabilidades de que aparezcan los genes A, B y 0, las probabilidades de aparición de los grupos de sangre corresponderán a las siguientes magnitudes:

Tabla 1

l (número de grupo)	Grupo	Combinaciones que dan este grupo	Probabilidades
1	0	00	r^2
2	A	AA, A0	$p^2 + 2pr$
3	B	BB, B0	$q^2 + 2qr$
4	AB	AB	$2pq$

Tabla 2

	l			
	1	2	3	4
$p_l(\theta)$	r^2	$p(p + 2r)$	$q(q + 2r)$	$2pq$
$\frac{\partial p_l(\theta)}{\partial p}$	$-2r$	$2r$	$-2q$	$2q$
$\frac{\partial p_l(\theta)}{\partial q}$	$-2r$	$-2p$	$2r$	$2p$

Sean ν_1, ν_2, ν_3, ν_4 las frecuencias de aparición de los grupos de sangre respectivos en la población sujeta a investigación, con un total de n personas. ¿Cómo hallar la e.v.m. par p y q? En nuestro caso las probabilidades $p_i(\theta)$, $\theta = (p, q)$ de aparición del i-ésimo grupo de sangre y sus derivadas parciales respecto a p y q se muestran en la tabla 2.

Por eso para la función logarítmica de verosimilitud $L(X, \theta) = \sum_{i=1}^{4} \nu_i \ln p_i(\theta)$ obtenemos

$$\begin{aligned} \frac{\partial L}{\partial p} &= \sum \frac{\nu_i}{p_i}\frac{\partial p_i}{\partial p} = -\frac{2\nu_1}{r} + \frac{2r\nu_2}{p(p+2r)} - \frac{2\nu_3}{q+2r} + \frac{\nu_4}{p}, \\ \frac{\partial L}{\partial q} &= \sum \frac{\nu_i}{p_i}\frac{\partial p_i}{\partial q} = -\frac{2\nu_1}{r} - \frac{2\nu_2}{p+2r} + \frac{2r\nu_3}{q(q+2r)} + \frac{\nu_4}{q}. \end{aligned} \tag{4}$$

Igualando a cero estas derivadas, llegaremos al sistema de dos ecuaciones para θ^* de cuarto orden. La resolución de tal sistema presenta dificultades técnicas. Por eso es más simple hacer uso del teorema 1. Para esto notemos que son válidas las igualdades

$$p_1 = r^2, \quad p_1 + p_2 = (p+r)^2, \quad p_1 + p_3 = (q+r)^2. \tag{5}$$

Las estimaciones eficientes para p_i son iguales a $p_i^* = \nu_i/n$. Sustituyendo en (5) estas estimaciones y resolviendo las ecuaciones obtenidas, tenemos

$$p^* = \sqrt{p_1^* + p_2^*} - \sqrt{p_1^*}, \quad q^* = \sqrt{p_1^* + p_3^*} - \sqrt{p_1^*}.$$

Como p_i^* es la estimación asintóticamente normal de p_i (o sea, $(p_i^* - p_i)\sqrt{n} \in \Phi_{0,\, p_i(1-p_i)}$), en virtud de los teoremas del § 1.5, p^* y q^* también serán las estimaciones asintóticamente normales para p y q.

Para valerse del teorema 1 sólo queda calcular la matriz $(L''(X, \theta^*))^{-1}$ o matriz $(nI(\theta^*))^{-1}$, $\theta^* = (p^*, q^*)$.

Citemos el ejemplo de una muestra real X obtenida como resultado del examen de $n = 353$ personas.

La distribución de la gente por grupos de sangre se da en la tabla 3.

Tabla 3

	0	*A*	*B*	*AB*	*Total*
ν_i	121	120	79	33	353
p_i^*	0,343	0,340	0,224	0,093	1

Tabla 3A

	0	*A*	*B*	*AB*
$p_i(\theta^*)$	0,351	0,343	0,226	0,080
$p_i(\theta_1^*)$	0,337	0,347	0,231	0,085

De esta tabla se deduce $p^* = 0{,}241$, $q^* = 0{,}167$, $r^* = 1 - p^* - q^* = 0{,}592$. Con ayuda de la tabla 2, para los elementos de la matriz $I(\theta)$, cuando

$\theta = \theta^*$; obtenemos

$$\sum\left(\frac{\partial p_i(\theta)}{\partial p}\right)^2\frac{1}{p_i(\theta)} = 4 + \frac{4r^2}{p(p+2r)} + \frac{4q}{q+2r} + \frac{2q}{p} = 9{,}970,$$

$$\sum\left(\frac{\partial p_i(\theta)}{\partial p}\right)^2\frac{1}{p_i(\theta)} = 4 + \frac{4p}{p+2r} + \frac{4r^2}{q(q+2r)} + \frac{2p}{q} = 13{,}761,$$

$$\sum\frac{\partial p_i(\theta)}{\partial p}\cdot\frac{\partial p_i(\theta)}{\partial q}\cdot\frac{1}{p_i(\theta)} = 4 - \frac{4r}{p+2r} - \frac{4r}{q+2r} + 2 = 2{,}585.$$

De aquí hallamos $|I(\theta^*)| = 130{,}512$.

$$I^{-1}(\theta^*) = \begin{vmatrix} 0{,}105 & -0{,}020 \\ -0{,}020 & 0{,}076 \end{vmatrix}.$$

De las fórmulas para $\frac{\partial L}{\partial p}$ y $\frac{\partial L}{\partial q}$ (véase (4)) obtenemos

$$L'(\theta^*, X) = (25{,}443,\ 34{,}161), \tag{6}$$

así que para la segunda aproximación de θ_1^* tenemos

$$\theta_1^* = \theta^* + \frac{1}{n}L'(\theta^*, X)I^{-1}(\theta^*) = (0{,}246,\ 0{,}173). \tag{7}$$

Esto nos da, para completar la tabla 3, las estimaciones expuestas en la tabla 3A.

La aplicación de una iteración más, en forma de (7), ya no modifica la estimación θ_1^* (dentro de los límites de la exactitud que utilizamos), ya que

$$L'(\theta_1^*, X) = (-0{,}076,\ -0{,}167)$$

(compárese con (6)), así que la tercera aproximación para $\hat{\theta}^*$ y todas las aproximaciones siguientes coincidirán con θ_1^*.

§ 27*. Propiedades de las estimaciones de verosimilitud máxima al faltar las condiciones de regularidad. Conciliabilidad

Este párrafo, al igual que el § 22, no entra en el curso principal de exposición y está dedicado al estudio de un caso irregular. Aquí nos limitaremos a demostrar la conciliabilidad fuerte de la e.v.m. en condiciones muy débiles respecto a $f_t(x)$, las cuales no suponen el cumplimiento de las condiciones (RR) o (R). Un estudio más detallado de las propiedades de la e.v.m. y de la relación de verosimilitud en el caso irregular véase en [48].

En todo el párrafo supondremos que se cumplen las condiciones (A_n),

(A_c) y (A_0) y designaremos la distancia de Kullback-Leibler $\varrho_1(\mathbf{P}_\theta, \mathbf{P}_t)$ por

$$\varrho(\theta, t) = \int \ln \frac{f_\theta(x)}{f_t(x)} \cdot f_\theta(x)\mu(dx).$$

Sabemos que $\varrho(\theta, t) > 0$ para $t \neq \theta$ si se cumple la condición (A_0).

Evidentemente, la condición (A_0) es necesaria para la conciliabilidad de la e.v.m., o sea, para la convergencia de $\hat{\theta}^* \underset{P_\theta}{\rightarrow} \theta$. Si, por ejemplo, $\varrho(\theta, t_0) = 0$ cuando $t_0 \neq \theta$, entonces los puntos θ y t_0 serán simplemente indistinguibles, las distribuciones $\mathbf{P}_\theta$ y $\mathbf{P}_{t_0}$ coincidirán y cualquiera que sea el lugar de convergencia de la e.v.m. $\hat{\theta}^*$, ésta no podrá ser conciliable si $X \in \mathbf{P}_\theta$ o si $X \in \mathbf{P}_{t_0}$.

La siguiente variante de la condición (A_0) se puede llamar uniforme (θ ha sido registrado):

$(\overline{A}_0)$ *Para cualquier* $\delta = \varepsilon(\delta) > 0$

$$\inf_{t:\, |t-\theta| \geq \delta} \varrho(\theta, t) > \varepsilon$$

con cierto $\varepsilon > 0$.

Es evidente que $(\overline{A}_0)$ será el corolario de (A_0), (A_c) y de la continuidad de $\varrho(\theta, t)$. Por consiguiente, en estas condiciones, la condición $(\overline{A}_0)$ también será necesaria.

Examinemos ahora la siguiente amplificación de la condición $(\overline{A}_0)$. Designemos

$$f_t^\Delta(x) = \sup_{|u| \leq \Delta} f_{t+u}(x).$$

(A_0^Δ). *Para cualquier* $\delta > 0$ *existe* $\Delta = \Delta(\delta) > 0$ *tal, que para todos* t, $|t - \theta| > \delta$,

$$\int \ln \frac{f_t^\Delta(x)}{f_\theta(x)} \cdot f_\theta(x)\mu(dx) < -\varepsilon \tag{1}$$

con cierto $\varepsilon > 0$.

Esta condición resulta suficiente para la conciliabilidad fuerte de la e.v.m. La misma es parecida a la condición $(\overline{A}_0)$ y en este sentido se asemeja a la condición necesaria. Una sola condición $(\overline{A}_0)$ no es suficiente para la conciliabilidad de la e.v.m. (véase la observación 1).

Teorema 1. *Si se cumple la condición* (A_0^Δ), *entonces la e.v.m.* $\hat{\theta}^*$ *es fuertemente conciliable.*

Demostración. La e.v.m. $\hat{\theta}^*$ es el punto t en el que se alcanza el máximo de la función $\psi(t, \theta, \mathbf{P}_n^*)$, donde

$$\psi(\theta, t, \mathbf{P}) = \int \ln \frac{f_t(x)}{f_\theta(x)} \mathbf{P}(dx).$$

Como $\psi(\theta, \hat{\theta}^*, \mathbf{P}_n^*) \geqslant \psi(\theta, \theta, \mathbf{P}_n^*) = 0$, para demostrar el teorema es suficiente convencerse de que con $\mathbf{P}_\theta$-probabilidad igual a 1,

$$\lim_{n\to\infty} \sup \sup_{|t-\theta|\geqslant\delta} \psi(\theta, t, \mathbf{P}_n^*) < -\varepsilon$$

con cierto $\varepsilon > 0$. (Esto precisamente significará que para c.t. $X_\infty \in \mathbf{P}_\theta$, a partir de cierto $n = n(X_\infty) < \infty$, se cumple $|\hat{\theta}^* - \theta| < \delta$). Supongamos que se ha registrado δ y que Δ satisface la condición (1). Recubramos el conjunto $\Theta \setminus [\theta - \delta, \quad \theta + \delta]$ con segmentos $\Delta_k = \{t: \quad |t - t_k| \leqslant \Delta\}$, $k = 1, \ldots, N < \infty$, donde $t_k \in \Theta$, $t_k \notin [\theta - \delta, \theta + \delta]$. En este caso, según la ley fuerte de los grandes números,

$$\sup_{|t-\theta|\geqslant\delta} \psi(\theta, t, \mathbf{P}_n^*) \leqslant \max_k \sup_{t\in\Delta_k} \psi(\theta, t, \mathbf{P}_n^*) \leqslant$$

$$\leqslant \max_k \frac{1}{n} \sum_{i=1}^{n} \sup_{t\in\Delta_k} \ln \frac{f_t(x_i)}{f_\theta(x_i)} \xrightarrow[\text{c.s.}]{} \max_k \mathbf{M}_\theta \ln \frac{f_{t_k}^\Delta(x_1)}{f_\theta(x_1)} < -\varepsilon. \quad \triangleleft$$

Observación 1. Como ya hemos señalado, una sola condición $(\overline{A}_0)$ no es suficiente para la conciliabilidad de $\hat{\theta}^*$. Para convencerse de esto examinemos el ejemplo siguiente. Sea $\Theta = [0, 1]$, $\mathbf{P}_\theta = \mathrm{U}_{\theta, 1+\theta}$ cuando $0 \leqslant \theta \leqslant 1/2$ y cuando $\theta = 1$. Cuando $1 > \theta > 1/2$, la distribución $\mathbf{P}_\theta$ tiene una densidad de $f_\theta(x) = 1/\theta$ cuando $1 - \theta < x \leqslant 1$. Supongamos ahora que $X \in \mathbf{P}_0 = \mathrm{U}_{0,1}$. En este caso la condición $(\overline{A}_0)$ se cumple, ya que $\varrho(0, t) = -\infty$ cuando $t \neq 0$. Al mismo tiempo es fácil ver que $f_t(X) > 1$ cuando $t \in (1 - x_{(1)}, 1)$ y que $\hat{\theta}^* = 1 - x_{(1)} \xrightarrow[\text{c.s.}]{} 1$.

Las condiciones (A_0^δ) pueden ser representadas de manera equivalente en una forma algo distinta. Designemos $f_t^0(x) = \lim_{u\to t} \sup f_u(x)$.

Teorema 2. *La condición* (A_0^δ) *es equivalente al cumplimiento simultáneo de las dos condiciones siguientes*

(A_δ^0). *Para todos* $t \neq \theta$

$$\int \ln \frac{f_t^0(x)}{f_\theta(x)} \cdot f_\theta(x)\mu(dx) < 0.$$

(J). *Para todos* t *y cierto* $\Delta > 0$

$$\int \ln \frac{f_t^\Delta(x)}{f_\theta(x)} \cdot f_\theta(x)\mu(dx) < \infty.$$

La condición (J), al igual que (A_δ^0), (A_0^δ), significa la integrabilidad de las partes positivas de las funciones subintegrales. Tales funciones es natural llamarlas *integrables superiormente*.

En virtud de (A_c), la condición (J) es, de hecho, equivalente a la limitación superior de la integral

$$\int \ln \frac{f^{\Theta}(x)}{f_\theta(x)} \cdot f_\theta(x)\mu(dx) < \infty, \tag{2}$$

donde $f^{\Theta}(x) = \sup_{t \in \Theta} f_t(x)$.

Demostración del teorema 2. El hecho de que de $(A\hat{\delta})$ resulte $(A\delta)$ y (J) es evidente. Ahora supongamos que se cumplen $(A\delta)$ y (J). Si admitimos que $((A\hat{\delta})$ no tiene lugar, existirán sucesiones $t_k \to t \in \Theta$, $\Delta_k \to 0$, $\varepsilon_k \to 0$ tales, que

$$\int \ln \frac{f^{\Delta}_{t_k}(x)}{f_\theta(x)} \cdot f_\theta(x)\mu(dx) > -\varepsilon_k.$$

Aquí la función subintegral es mayorada, en virtud de la condición (J), por la función superiormente integrable, por eso, en virtud del lema de Fatou,

$$\lim_{k\to\infty}\sup \int \ln \frac{f^{\Delta_k}_{t_k}(x)}{f_\theta(x)} f_\theta(x)\mu(dx) \leqslant \int \ln \frac{f_t^{\circ}(x)}{f_\theta(x)} f_\theta(x)\mu(dx) < 0.$$

Hemos obtenido la contradicción que demuestra el teorema. ◁

Ahora expondremos unas condiciones bastante más simples, que demuestran el cumplimiento de $(A\delta)$ y (J) y, por lo tanto, la conciliabilidad fuerte de la e.v.m.

Definición 1. Diremos que $f_t(x)$ pertenece a la clase D_0, si para cada $t \in \Theta$ existe un conjunto $C_t \in \mathfrak{B}_{\mathscr{X}}$, $\mathbf{P}_\theta(C_t) = 1$ en el que $f_t(x)$ es continua respecto a t: $f_{t_k}(x) \to \bar{f}_t(x)$ cuando $t_k \to t$, $x \in C_t$.

Además de las $f_t(x)$ continuas (respecto a t) en el conjunto C, $\mathbf{P}_\theta(C) = 1$ independiente de t, a la clase D_0 también pertenecen, por supuesto, otras funciones, tales, por ejemplo, para las cuales $f_t(x)$ en el plano (t, x) tiene líneas de discontinuidad aisladas y desprovistas de partes paralelas al eje x. Así será, en particular, si $f_t(x)$, como función de x, tiene discontinuidades aisladas en los puntos $x_t^{(1)}$, $x_t^{(2)}$, ..., que dependen continuamente de t.

Teorema 3. *Si $f_t(x) \in D_0$ y se cumple (J), entonces también se cumple la condición $(A\delta)$ y, por lo tanto, la e.v.m. $\hat{\theta}^*$ es fuertemente conciliable.*

Demostración. Si $f_t(x) \in D_0$, entonces $f_t^{\circ}(x) = f_t(x)$ cuando $x \in C_t$ y, por lo tanto,

$$\int \ln \frac{f_t^{\circ}(x)}{f_\theta(x)} f_\theta(x)\mu(dx) = -\varrho(\theta, t) < 0. \quad ◁$$

Corolario 1. *Si $f_t(x) \in D_0$ está limitada, y la integral*

$$\int f_\theta(x) \ln f_\theta(x)\mu(dx) \tag{3}$$

es finita, la e.v.m. es fuertemente conciliable.

La afirmación del corolario 1 se deduce directamente del teorema 3, ya que el carácter limitado de $f_t(x)$ y la finitud de la integral (3) conducen a (J).

Corolario 2. *Si*

$$\varphi(\Delta) = \int \sup_{|u|<\Delta} |f_{t+u}(x) - f_t(x)|\mu(dx) \to 0 \tag{4}$$

cuando $\Delta \to 0$, la e.v.m. es fuertemente conciliable.

Demostración. Hagamos uso del teorema 3. La pertenencia de $f_t(x) \in D_0$ es evidente, ya que (4) puede cumplirse tan sólo en el caso en que $f_{t+u}(x) \to f_t(x)$ cuando $u \to 0$ para c.t. $[\mu]$ valores de x.

Luego,

$$\int f_t^\Delta(x)\mu(dx) \leqslant \varphi(\Delta) + \int f_t(x)\mu(dx) = \varphi(\Delta) + 1,$$

y la condición (4) también significa la integrabilidad de $f_t^\Delta(x)$. Como $\ln \dfrac{f_t^\Delta(x)}{f_\theta(x)} \leqslant \dfrac{f_t^\Delta(x)}{f_\theta(x)} - 1$, de aquí obtenemos que la integral en las condiciones (J) no supera

$$\int f_t^\Delta(x)\mu(dx) - 1 \leqslant \varphi(\Delta). \quad \triangleleft$$

En vez de (4) podríamos exigir la convergencia a cero de la magnitud

$$\varphi_1(\Delta) = \int \sup_{|u|\leqslant\Delta} (\sqrt{f_{t+u}(x)} - \sqrt{f_t(x)})^2\mu(dx),$$

ya que $\varphi(\Delta)$ se puede estimar con ayuda de $\varphi_1(\Delta)$ utilizando la desigualdad

$$\varphi(\Delta) \leqslant \int \sup_{|u|\leqslant\Delta} |\sqrt{f_{t+u}(x)} - \sqrt{f_t(x)}| \sup_{|u|\leqslant\Delta} |\sqrt{f_{t+u}(x)} + \sqrt{f_t(x)}|\mu(dx) \leqslant$$

$$\leqslant \varphi_1^{1/2}(\Delta)\left[\int \sup_{|u|\leqslant\Delta} (\sqrt{f_{t+u}(x)} - \sqrt{f_t(x)} + 2\sqrt{f_t(x)})^2\mu(dx)\right]^{1/2} \leqslant$$

$$\leqslant [2\varphi_1(\Delta)(\varphi_1(\Delta) + 4]^{1/2}.$$

Corolario 3. *Si $f_t(x)$ es derivable respecto a t para c.t. $[\mu]$ valores de x, y*

$$\int |f_t'(x)|\mu(dx) < c < \infty, \tag{5}$$

entonces la e.v.m. θ^* es fuertemente conciliable. La condición (5) siempre se cumple si la información de Fisher $I(t)$ está limitada.

Aquí hemos llegado al mismo resultado que podríamos obtener del teorema 23.2. El método de demostración de este último (véanse los §§ 21, 23) muestra que el carácter limitado de $I(t)$ o (5) no son esenciales para la afirmación del corolario 3 si la distancia de Hellinger $\varrho_3(\mathbf{P}_\theta, \mathbf{P}_{\theta+\Delta})$ está uniformemente separada del cero cuando $|\Delta| \geqslant \delta > 0$.

Demostración. La pertenencia de $f_t(x) \in D_0$ es evidente. Para el cumplimiento de la condición (J) es suficiente, como hemos visto en la demostración del corolario 2, la integrabilidad de $f_t^\Delta(x)$. Pero

$$\int f_t^\Delta(x)\mu(dx) \leqslant \int \left[f_t(x) + \int_{-\Delta}^{\Delta} |f'_{t+u}(x)| du \right] \mu(dx) =$$
$$= 1 + \int_{-\Delta}^{\Delta} \left[\int |f'_{t+u}(x)| \mu(dx) \right] du \leqslant 1 + 2\Delta c.$$

Queda hacer uso del teorema 3. La última afirmación del corolario 3 se deduce de la desigualdad de Cauchy — Buniakovski, ya que, en virtud de esta desigualdad, $\int |f'_t(x)| \mu(dx) \leqslant I^{1/2}(t)$. ◁

Corolario 4. *Sea θ el parámetro de desplazamiento de la familia $f_\theta(x) = f(x - \theta)$, $\int f(x) \ln f(x) dx > -\infty$. Si la función $f(x)$ está limitada (de lo contrario el método de verosimilitud máxima pierde su sentido (véase el § 26)) y tiene un conjunto B de puntos de discontinuidad, cuya medida de Lebesgue de clausura $\mu(B^c)$ es igual a cero, entonces la e.v.m. $\hat{\theta}^*$ es fuertemente conciliable.*

Demostración. Verifiquemos el cumplimiento de las condiciones del teorema 3. La condición (J) se cumple de modo evidente. La pertenencia de $f_t(x) \in D_0$ se desprende de la definición de D_0 en que es necesario poner $C_t = \overline{B^c} - t$ (este es el desplazamiento del conjunto $\overline{B^c}$ en t, y $\overline{B^c}$ es la adición a la clausura del conjunto B). En vista de que el conjunto $\overline{B^c}$ está abierto, $x - t \in \overline{B^c} - t$ conduce a $x - t_k \in \overline{B^c} - t$ para $|t_k - t|$ bastante pequeñas. Esto quiere decir que $f(x - t_k) \to f(x - t)$. El corolario queda demostrado.

Cabe señalar que en las condiciones del corolario 4 es inútil suponer que se ha cumplido la condición (A_0), puesto que ésta se cumple automáticamente. Si admitamos que (A_0^δ) no tiene lugar, llegaremos a la periodicidad de la función $f(x)$, lo que es imposible.

En cuanto a las condiciones del corolario 4, señalaremos que la condición de "continuidad" de $f(x)$, enunciada en este corolario, es muy débil. Pero, por lo visto, tampoco esta condición es esencial. Lo confirma, en cierta medida, el ejemplo siguiente.

Ejemplo 1. Sea $f(x)$ una función arbitraria que tiene un portador limitado
$(a, b) = \{x: f(x) > 0\}$. Entonces

$$\mathbf{P}_\theta(|\hat{\theta}^* - \theta| > \delta) \leqslant (1 - F_0(a + \delta))^n + F_0^n(b - \delta), \quad (6)$$

donde $F_\theta(x) = \int_{-\infty}^{x} f_\theta(y)dy$. La desigualdad (6) significa la conciliabilidad fuerte de $\hat{\theta}^*$. Esto se deduce de las relaciones que tienen la forma siguiente:

$$\{\hat{\theta}^* - \theta > \delta\} \subset \left\{ \prod_{i=1}^{n} f_{\theta+\delta}(x_i) > 0 \right\} \subset \bigcap_{i=1}^{n} \{x_i \geqslant a + \theta + \delta\},$$

$$\mathbf{P}_\theta(\hat{\theta}^* - \theta > \delta) \leqslant [1 - F_\theta(a + \theta + \delta)]^n = [1 - F_0(a + \delta)]^n.$$

Desde cierto punto de vista la condición de finitud de la integral $\int f(x) \ln f(x)\, dx$ en el corolario 4 tampoco es esencial: se puede construir fácilmente un ejemplo cuando esta integral se convierte en $-\infty$ y la condición (J) queda cumplida.

De las observaciones del § 2.18 se desprende que todo lo dicho en el corolario 4 y después de éste conserva por completo su validez para el parámetro de escala.

§ 28. Resultados de los §§ 23—27 para el caso del parámetro multidimensional

En este párrafo trasladaremos al caso multidimensional todos los resultados principales de los §§ 23—27. Dichos resultados serán expuestos en el mismo orden que en los párrafos indicados, con la particularidad de que sólo nos detendremos en los momentos donde el carácter multidimensional modifica la formulación del resultado o exige la modificación de los razonamientos.

Así pues, supongamos $\theta \in \Theta \subset R^k$, $k > 1$. Las enunciaciones de las condiciones (A_μ), (A_c) y (A_0), al igual que las definiciones de la relación de verosimilitud

$$Z(u) = \frac{f_{\theta+u}(X)}{f_\theta(X)}$$

y la distancia de Hellinger

$$r(u) = \varrho(\mathbf{P}_{\theta+u}, \mathbf{P}_\theta) = \int (\sqrt{f_{\theta+u}(x)} - \sqrt{f_\theta(x)})^2 \mu(dx),$$

no están relacionadas de ningún modo con la dimensión.

1. Desigualdades para la relación de verosimilitud (resultados del § 23). Para estudiar el comportamiento de la función $Z(u)$ en el entorno del cero

necesitaremos la condición siguiente: *la función* $\sqrt{f_\theta(x)}$ *es derivable respecto a* θ, *y la matriz de información de Fisher*

$$I(\theta) = |I_{ij}(\theta)| = \left\| \mathbf{M}_\theta \frac{\partial}{\partial\theta_i} l(x_1, \theta) \frac{\partial}{\partial\theta_j} l(x_1, \theta) \right\|, \qquad (1)$$

para todos $\theta \in \Theta$, *está limitada y definida positivamente.*

Dada esta condición, del teorema 21.3A resulta que para todos θ,

$$0 < g \leqslant \frac{r(u)}{|u|^2} \leqslant h = \frac{1}{4} \sup_\theta \operatorname{Sp} I(\theta) < \infty. \qquad (2)$$

Aquí y en lo sucesivo $|u|$ significa la norma euclídea $|u| = \sqrt{u_1^2 + \ldots + u_k^2}$ del vector $u = (u_1, \ldots, u_k)$.

La primera afirmación del teorema 23.1 y su demostración se trasladan al caso multidimensional sin camios algunos, ya que, de hecho, las mismas no están relacionadas con la dimensión.

Teorema 1. *Si se cumple* (2), *entonces*

$$\mathbf{M}_\theta Z^{1/2}(u) \leqslant e^{-ng|u|^{2}/2}.$$

Para generalizar el teorema 23.2 necesitaremos una condición adicional que consiste en que

$$\gamma = \sup_\theta \mathbf{M}_\theta |l'(x_1, \theta)|^s < \infty \qquad (3)$$

con cierto $s > k$.

Teorema 2 (análogo del teorema 23.2). *Si se cumplen las condiciones* (2) y (3), *entonces, con todos* z, $n \geqslant 1$

$$\mathbf{P}_\theta\left(\sup_{|v| > u} Z\left(\frac{v}{\sqrt{n}}\right) > e^z \right) \leqslant c\gamma e^{-z} + e^{-z/2} e^{-\beta u^2}, \qquad (4)$$

donde $c < \infty$, $\beta > 0$ sólo dependen de k, g y s.

Para demostrar esta afirmación, en el caso unidimensional hemos utilizado la posibilidad de estimar $\sup_{u \in (0, 1)} p(u)$ por los valores de $p(0)$ y $\int_0^1 |p'(u)| du$. En el caso multidimensional, tal enfoque choca con dificultades, puesto que el valor máximo de $p(u)$ en cierta región DCR^k, $k > 1$, no puede ser estimado, hablando en general, por los valores de $p(u_0)$, $u_0 \in D$, y la integral de $p'(u)$ ($p'(u) = \operatorname{grad} p(u)$), por una curva registrada cualquiera de D. Existen, por lo menos, dos vías para superar esta dificultad.

La primera es absolutamente análoga al enfoque unidimensional y consiste en utilizar la estimación que tiene la siguiente forma (en esta fórmula,

para simplificar la escritura, nos limitamos al caso bidimensional $k = 2$):

$$\sup_{u\in K_{0,1}} p(u) \leqslant |p(0)| + \int_0^1 \left|\frac{\partial p((0,\ u_2))}{\partial u_2}\right| du_2 + \int_0^1 \left|\frac{\partial p((u_1,\ 0))}{\partial u_1}\right| \times$$

$$\times\ du_1 + \int_0^1\int_0^1 \left|\frac{\partial^2 p(u)}{\partial u_1 \partial u_2}\right| du_1\, du_2,$$

donde $u = (u_1,\ u_2)$, $K_{0,1}$ es el cubo unitario $K_{0,1} = \{u\colon\ 0 \leqslant u_j \leqslant 1\ j = 1,\ \ldots,\ k\}$. Sin embargo, para utilizar este enfoque debemos suponer que existen derivadas de k-ésimo orden de la función $l_\theta(x)$ ($f_\theta(x)$) (véase la definición de la función p en el párrafo 23) y saber apreciar los valores medios (que necesitamos) de las derivadas de la función p del l-ésimo orden, $l \leqslant k$.

La segunda vía es más conveniente, ya que utiliza la posibilidad de estimar $\sup_{u\in K_{0,1}} p(u)$ a través de los valores de $p(0)$ y

$$\int_{K_{0,1}} |p'(u)|^s du \quad (p'(u) = \text{grad}\ p(u),\ u = (u_1,\ \ldots,\ u_k))$$

con cierto $s > k$ (cuando $s = k$ la estimación es imposible). En este caso, sin duda, debemos disponer de las estimaciones para $\mathbf{M}_\theta |p'(u)|^s$ cuando $s > k$. La obtención de todas las estimaciones aquí necesarias presenta ciertas dificultades y requiere mucho espacio. Por eso la demostración del teorema 2 para el caso multidimensional se da en el Suplemento VII.

También debemos señalar que en el libro editado en ruso se utilizó otro método de demostración del teorema 2 (véanse las observaciones bibliográficas referentes al Suplemento VII).

Las demostraciones de las afirmaciones acerca de la conciliabilidad de la e.v.m. y acerca de las estimaciones para los momentos en el punto 2 del § 23, no están relacionadas con la dimensión. Las propias afirmaciones se conservarán en la forma siguiente.

Teorema 3 (análogo del teorema 23.3). *Si se cumplen las condiciones* (2) *y* (4), *entonces para cualesquiera* z, $n \geqslant 1$ *es válida* (23.6) *sustituyendo el número* $g/4$ *por* β (véase el teorema 2).

Las afirmaciones de los corolarios 23.1 y 23.2 conservan por completo su validez sustituyendo igualmente $g/4$ por β.

2. Propiedades asintóticas de la relación de verosimilitud (resultados del § 24).

En el caso multidimensional, por condiciones (*RR*) entenderemos el conjunto de condiciones siguientes:

1) *Condiciones* (A_0), (A_c), (*R*).

2) *Derivabilidad continua de segundo orden respecto a θ dentro de Θ, de la función $l(x, t)$ para c.t. $[\mu]$ valores de x. En este caso se supone que las derivadas*

$$l_{ij}''(x, t) = \frac{\partial^2 l(x, t)}{\partial t_i \partial t_j}$$

admiten la mayorante $l(x)$ que no depende de t: $|l_{ij}''(x, t)| \leqslant l(x)$, para la cual

$$\mathbf{M}_t l(\mathbf{x}_1) = \int l(x) f_t(x) \mu(dx)$$

*converge uniformemente**) *en $t \in \Theta$.*

3) *Además, supondremos, siempre que sea necesario, que se cumple la condición* (3).

Al igual que en el caso unidimensional, necesitaremos las dos propiedades siguientes que se deducen de (RR):

1) *Posibilidad de derivar dos veces respecto a θ bajo el signo integral en la igualdad*

$$\int f_\theta(x) \mu(dx) = 1,$$

que significa la validez de las relaciones

$$\int \frac{\partial}{\partial \theta_i} f_\theta(x) \mu(dx) = 0, \qquad \int \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_\theta(x) \mu(dx) = 0. \tag{5}$$

2) *Convergencia uniforme de la integral $I(\theta)$:*

$$\sup_\theta \mathbf{M}_\theta [(l'(\mathbf{x}_1, \theta))^2; \quad |l'(\mathbf{x}_1, \theta)| > N] \to 0 \tag{6}$$

cuando $N \to \infty$.

Estas propiedades se demuestran en el Suplemento VI. Para simplificar la exposición, las referidas propiedades pueden ser intoducidas en las condiciones (RR).

En virtud de las igualdades

$$l_i'(x, \theta) = \frac{1}{f_\theta(x)} \cdot \frac{\partial f_\theta(x)}{\partial \theta_i},$$

$$l_{ij}''(x, \theta) = \frac{1}{f_\theta(x)} \frac{\partial^2 f_\theta(x)}{\partial \theta_i \partial \theta_j} - \frac{1}{f_\theta^2(x)} \cdot \frac{\partial f_\theta(x)}{\partial \theta_i} \cdot \frac{\partial f_\theta(x)}{\partial \theta_j},$$

de las relaciones (5) resulta que

$$\mathbf{M}_\theta l_i'(\mathbf{x}_1, \theta) = 0,$$

$$\mathbf{M}_\theta l_{ij}''(\mathbf{x}_1, \theta) = -\mathbf{M}_\theta l_i'(\mathbf{x}_1, \theta) l_j'(\mathbf{x}_1, \theta) = -I_{ij}(\theta).$$

*) Véase la nota en la pág. 226, acerca de la convergencia uniforme en el § 24.

Al igual que en el caso unidimensional, las condiciones (RR) significan que tendrán lugar las afirmaciones de los teoremas del § 23 acerca de las estimaciones para

$$\sup_{|v|\geqslant u} Z(v/\sqrt{n}) \text{ y para } \sqrt{n}(\hat{\theta}^* - \theta).$$

Al cumplirse las condiciones (RR), también serán válidos los siguientes análogos de los lemas 24.1 y 24.2.

Lema 1. *Las funciones* $l_{ij}''(x, \theta)$ son continuas "*por término medio*":

$$\mathbf{M}_\theta \omega_\Delta''(\mathrm{x}_1) \to 0$$

es uniforme respecto a θ *cuando* $\Delta \to 0$, *donde* $\omega_\Delta''(x) = \max\limits_{i,j} \sup\limits_{\theta, |u|<\Delta} |l_{ij}'' \times$ $\times (x, \theta + u) - l_{ij}''(x, \theta)|$.

La demostración repite exactamente los razonamientos del lema 24.1. ◁

Pongamos

$$\gamma_n(\delta, \theta) = \sup_{\substack{\Delta\leqslant\delta\\ |\omega|=1}} \left| \frac{(L'(X, \theta + \omega\Delta), \omega) - (L'(X, \theta), \omega)}{n\Delta} + \omega I(\theta)\omega^T \right|.$$

Lema 2. (análogo del lema 24.2). *Supongamos que se cumplen las condiciones (RR) y que* $\delta_n > 0$ *es cualquier sucesión que converge a cero. Entonces, para* $X \in \mathbf{P}_\theta$

$$\gamma_n(\delta_n, \theta) \underset{c.s.}{\to} 0, \quad \gamma_n(\delta_n, \hat{\theta}^*) \underset{c.s.}{\to} 0.$$

En estas relaciones, los valores de $I(\theta)$ *e* $I(\hat{\theta}^*)$ *pueden sustituirse uno por otro.*

Demostración. Al igual que en el caso unidimensional, es suficiente convencerse de que $\gamma_n(\delta_n) \to 0$, donde

$$\gamma_n(\delta) = \sup_{\substack{\Delta\leqslant\delta\\ |\omega|=1}} \left| \frac{(L'(X, \theta + \omega\Delta), \omega) - (L'(X, \theta), u)}{n\Delta} - \frac{\omega L''(X, \theta)\omega^T}{n} \right|.$$

Pero $\gamma_n(\delta_n) \leqslant \frac{1}{n} \sum\limits_i \sum\limits_{k,j} \omega_{\delta_n}''(\mathrm{x}_i)|\omega_k\omega_j|$, donde $\omega_\delta''(x)$ es el módulo máximo de continuidad de las funciones $l_{ij}''(x, \theta)$. Como

$$\sum_{k,j} |\omega_k\omega_j| \leqslant k|\omega|^2 = k,$$

entonces

$$\gamma_n(\delta_n) \leqslant \frac{k}{n} \sum_i \omega_{\delta_n}''(\mathrm{x}_i). \tag{7}$$

La demostración ulterior se base en el lema 1 y repite exactamente los razonamientos del lema 24.2. ◁

La generalización del teorema 24.1 para el caso multidimensional aquí es el

Teorema 4. *Supongamos que se cumplen las condiciones (RR) y que $\delta_n > 0$, $n = 1, 2, \ldots$, es cualquier sucesión convergente a cero. En este caso, si $X \in \mathbf{P}_\theta$, para u tales, que $|u/\sqrt{n}| \leqslant \delta_n$,*

$$Y(u) \equiv \ln Z(u/\sqrt{n}) = (\xi_n, u) - \frac{1}{2} uI(\theta)u^T(1 + \varepsilon_n(X, \theta, u)), \quad (8)$$

donde $|\varepsilon_n(X, \theta, u)| \leqslant \varepsilon_n(X, \theta) \underset{c.s.}{\to} 0$, $\xi_n = \frac{1}{\sqrt{n}} \operatorname{grad} L(X, \theta) = \frac{1}{\sqrt{n}} L'(X, \theta) \in$ $\in \Phi_{0, I(\theta)}$.

El valor de $u^* = \sqrt{n}(\hat{\theta}^* - \theta)$ con el que $Y(u)$ alcanza su valor máximo es representable en la forma

$$u^* = \xi_n I^{-1}(\theta)(E + \varepsilon_n(X, \theta)), \quad \varepsilon_n(X, \theta) \underset{c.s.}{\to} 0, \quad (9)$$

donde E es la matriz unidad. Además,

$$2Y(u^*) = \xi_n I^{-1}(\theta)\xi_n^T(1 + \varepsilon_n(X, \theta)) \in$$

$$\in \frac{1}{2} \xi I^{-1}(\theta)\xi^T \in H_k, \quad \xi \in \Phi_{0, I(\theta)}. \quad (10)$$

A la par con (8) es válida la representación

$$Y(u) - Y(u^*) = \frac{1}{2}(u - u^*)I(\theta)(u - u^*)^T(1 + \varepsilon_n(X, \theta, u)),$$

$$|\varepsilon_n(X, \theta, u)| \leqslant \varepsilon_n(X, \theta).$$

En todas las afirmaciones mencionadas se puede sustituir $I(\theta)$ por $I(\hat{\theta}^*)$.

Al igual que en el § 24, en este párrafo, por $\varepsilon_n(X, \theta)$ entendemos las distintas sucesiones que poseen la propiedad de $\varepsilon_n(X, \theta) \underset{c.s.}{\to} 0$ respecto a $\mathbf{P}_\theta$.

También debemos señalar que el miembro principal en (8) puede ser escrito de la forma siguiente:

$$\xi_n u^T - \frac{1}{2} uI(\theta)u^T =$$

$$= -\frac{1}{2}(u - \xi_n I^{-1}(\theta))I(\theta)(u - \xi_n I^{-1}(\theta))^T + \frac{1}{2} \xi_n I^{-1}(\theta)\xi_n^T.$$

Esto corresponde a la densidad de una distribución normal multidimensional con media $\xi_n I^{-1}(\theta)$ y con matriz de segundos momentos $I^{-1}(\theta)$.

La demostración del teorema 4 es completamente análoga a la del teorema 24.1. Del lema 2, cuando $\Delta \leqslant \delta_n$, obtenemos

$$(L'(X, \theta + \Delta\omega), \omega) = (L'(X, \theta), \omega) -$$
$$- n\Delta\omega I(\theta)\omega^T(1 + \varepsilon_n(X, \theta, \Delta\omega)), \quad |\varepsilon_n(X, \theta, \Delta\omega)| \leqslant \varepsilon_n(X, \theta).$$

Integrando esta igualdad respecto a Δ de 0 a $|u|/\sqrt{n}$ y poniendo $\omega = u/|u|$, obtenemos

$$L(X, \theta + u/\sqrt{n}) - L(X, \theta) = \int_0^{|u|/\sqrt{n}} (L'(X, \theta + \Delta u), u)d\Delta =$$
$$= \frac{|u|}{\sqrt{n}}(L'(X, \theta), \omega) - \frac{|u|^2}{2}\omega I(\theta)\omega^T(1 + \varepsilon_n(X, \theta, u)) =$$
$$= (\xi_n, u) - \frac{1}{2}uI(\theta)u^T(1 + \varepsilon_n(X, \theta, u)), \quad |\varepsilon_n(X, \theta, u)| \leqslant \varepsilon_n(X, \theta).$$

Aquí, según el teorema central multidimensional de límite (véase el suplemento V),

$$\xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l'(x_i, \theta) \in \Phi_{0, I(\theta)}.$$

La representación (8) queda demostrada. Las demás afirmaciones del teorema se demuestran absolutamente igual que en el teorema 24.1, teniendo en cuenta tan sólo las modificaciones de mostradas relacionadas con la multidimensión. La relación

$$\frac{1}{2}\xi I^{-1}(\theta)\xi^T \in \mathbf{H}_k$$

en (10) se deduce de las propiedades de la distribución normal (véase el punto 4 del § 2.2). ◁

Con arreglo a la relación (10) también es útil la siguiente

Observación 1. La matriz $I^{-1}(\theta)$, junto con $I(\theta)$, es positivamente definida, y existe una matriz $I^{-1/2}(\theta)$ que es la raíz cuadrada de $I^{-1}(\theta)$, o sea, una matriz que satisface la relación

$$I^{-1/2}(\theta)I^{-1/2}(\theta) = I^{-1}(\theta).$$

En efecto, si cierta matriz $\mathbf{M} > 0$ (está positivamente definida), entonces existe una matriz ortogonal C para la cual $CMC^T = \text{diag}(\lambda_1, ..., \lambda_k)$ es una matriz diagonal con elementos positivos $\lambda_i > 0$ en la diagonal. Si ponemos ahora $\mathbf{M}^{1/2} = C^T \text{diag}(\lambda_1^{1/2}, ..., \lambda_k^{1/2})C$, obtenemos, evidentemente, la raíz cuadrada de $\mathbf{M}$.

Valiéndonos de esto y de la simetría de la matriz $I^{-1}(\theta)$, podemos (10) escribir en la forma

$$\frac{1}{2}(\xi_n I^{-1/2}(\theta))(\xi_n I^{-1/2}(\theta))^T.$$

Aquí el vector $\eta_n = \xi_n I^{-1/2}(\theta)$ es, evidentemente, la suma normalizada de los vectores aleatorios igualmente distribuidos, con una media nula y una matriz de segundos momentos

$$\mathbf{M}_\theta(\xi_n I^{-1/2}(\theta))^T(\xi_n I^{-1/2}(\theta)) = \mathbf{M}_\theta I^{-1/2}(\theta)\xi_n^T\xi_n I^{-1/2}(\theta) = E,$$

puesto que

$$\mathbf{M}_\theta \xi_n^T \xi_n = \mathbf{M}_\theta(l'(x_1, \theta))^T(l'(x_1, \theta)) = I(\theta).$$

Esto significa que según el teorema central multidimensional del límite, $\xi_n I^{-1/2}(\theta) \in \Phi_{0,E}$.

Teorema 5 (análogo del teorema 24.2). *Supongamos que se cumplen las condiciones del teorema 24.2 para $\theta \in R^k$ multidimensional y para $\alpha = \beta/2$ (β está definido en el teorema 2). En este caso*

$$J = \int w(u^* - u)q(\theta + u/\sqrt{n})Z(u/\sqrt{n})\Pi(du) = e^{Y(u^*)}q(\theta) \times$$

$$\times \left[\int w(u^* - u) \exp\left\{-\frac{1}{2}(u - u^*)I(\theta)(u - u^*)^T\right\}\Pi(du) + \varepsilon_n(X, \theta)\right]. \tag{11}$$

Si Π es la medida de Lebesgue, y $\Pi(du) = du$, entonces

$$J = \frac{(2\pi)^{k/2}}{\sqrt{|I(\theta)|}} e^{Y(u^*)}q(\theta)(\mathbf{M}w(\eta) + \varepsilon_n(X, \theta)), \tag{12}$$

donde $\varepsilon_n(X, \theta) \underset{c.s.}{\to} 0$, $\eta \in \Phi_{0, I^{-1}(\theta)}$ (la sucesión $\varepsilon_n(X, \theta)$ es vectorial si $w(t)$ es una función vectorial).

El teorema 5 se demuestra igual que el teorema 24.2, puesto que la demostración de este último no está relacionada con la dimensión.

3. Propiedades de la e.v.m. (resultados del § 25). Aquí siempre supondremos que se cumplen las condiciones (RR).

El análogo del teorema 25.1 tendrá la forma siguiente.

Teorema 6. *La e.v.m. $\hat{\theta}^*$ es una estimación asintóticamente normal, con la particularidad de que la convergencia*

$$u^* = (\hat{\theta}^* - \theta)\sqrt{n} \in \Phi_{0, I^{-1}(\theta)}$$

tiene lugar junto con los momentos de cualquier orden. En particular,

$$\mathbf{M}_\theta n(\hat{\theta}^* - \theta)^T(\hat{\theta}^* - \theta) \to I^{-1}(\theta). \tag{13}$$

Además, para cualquier función continua $w(t)$ tal, que $|w(t)| < e^{\beta|t|^2/2}$ (el número β está definido en el teorema 2),

$$\mathbf{M}_\theta w(u^*) \to \mathbf{M}w(\eta), \quad \eta \in \Phi_{0, I^{-1}(\theta)}.$$

La relación (13) significa que $\hat{\theta}^* \in K_{\Phi,2}$.

La afirmación del teorema 6 se desprende del teorema 4 (véase (9)) y del análogo multidimensional de corolario 23.2 que se deduce del teorema 3 (compárese con la demostración del teorema 25.1). ◁

Definamos la clase $\tilde{K}_0$ como población de las estimaciones θ^* para las cuales el desplazamiento $b(\theta) = (b_1(\theta), ..., b_k(\theta)) = \mathbf{M}_\theta\theta^* - \theta$ posee las propiedades

$$|b(\theta)| = o(1/\sqrt{n}), \qquad b_{ij}(\theta) = \frac{\partial b_i(\theta)}{\partial \theta_j} \to 0$$

cuando $n \to \infty$.

El análogo de los teoremas 25.2 y 25.3 aquí tiene la misma forma.

Teorema 7. *$\hat{\theta}^*$ es una estimación asintóticamente R-eficiente. Además, $\hat{\theta}^* \in \tilde{K}_0$ también es asintóticamente eficiente en $\tilde{K}_0$.*

El carácter asintóticamente R-eficiente de $\hat{\theta}^*$, equivalente a (13), tiene lugar evidentemente. La pertenencia de $\hat{\theta}^* \in \tilde{K}_0$ y la eficacia asintótica en $\tilde{K}_0$ se demuestran completamente igual que en el caso unidimensional.

Pasemos ahora a examinar la propiedad del carácter asintóticamente bayesiano. El carácter asintóticamente R-bayesiano de la estimación θ^* significa, por definición, que (compárese con el § 20)

$$\mathbf{M}(\theta^* - \theta)^T(\theta^* - \theta) = J/n + o(1/n), \qquad J = \int I^{-1}(t)\mathbf{Q}(dt). \tag{14}$$

El carácter asintóticamente bayesiano de θ^* significa

$$\lim_{n\to\infty} \sup\, [nv(\theta^*) - nv(\theta^*_Q)] \leqslant 0, \tag{15}$$

donde θ^*_Q es la estimación bayesiana que minimiza $v(\theta^*) = \mathbf{M}(\theta^* - \theta) \times$ $\times\, V(\theta^* - \theta)^T$ para cualquier matriz V definida no negativamente.

Teorema 8 (análogo del teorema 25.4). *$\hat{\theta}^*$ es una estimación asintóticamente R-bayesiana. Si la distribución a priori $\mathbf{Q}$ tiene densidad respecto a la medida de Lebesgue en Θ, entonces $\hat{\theta}^*$ es una estimación asintóticamente bayesiana.*

La demostración es completamente análoga a la del teorema 25.4. La relación (14) para $\theta^* = \hat{\theta}^*$ se deduce del hecho de que

$$\lim_{n\to\infty} \mathbf{M}n(\hat{\theta}^* - \theta)^T(\hat{\theta}^* - \theta) =$$

$$= \mathbf{M} \lim_{n\to\infty} \mathbf{M}_\theta n(\hat{\theta}^* - \theta)^T(\hat{\theta}^* - \theta) = \mathbf{M}I^{-1}(\theta) = J.$$

El paso límite bajo el signo de la esperanza matemática (o sea, de la integral) aquí es legítimo, ya que la magnitud $\mathbf{M}_\theta n(\hat{\theta}^* - \theta)^T(\hat{\theta}^* - \theta)$ está limitada por una constante que no depende ni de n ni de θ (compárese con el corolario 23.2).

Para demostrar (15) notemos que, conforme al § 20, la desigualdad integral de Rao—Cramer, en el caso cuando **Q** tiene densidad, reviste el aspecto

$$\mathbf{M}n(\theta^* - \theta)^T(\theta^* - \theta) \geqslant J + o(1).$$

Esto significa que

$$nv(\theta^*_Q) \geqslant \sum v_{ij}J_{ij} + o(1),$$

donde $\|J_{ij}\| = J$, $\|v_{ij}\| = V$. Por otro lado, en virtud de (14) cuando $\theta^* = \hat{\theta}^*$,

$$nv(\hat{\theta}^*) = \sum v_{ij}J_{ij} + o(1).$$

Es evidente que de estas relaciones se deduce (15) cuando $\theta^* = \hat{\theta}^*$. ◁

Los análogos de los teoremas 25.5 y 25.6 también tendrán lugar. Por ejemplo, del teorema 5 se desprende

Teorema 9 (análogo del teorema 25.6). *Supongamos que $X \in \mathbf{P}_\theta$ y que θ es un punto interior arbitrario de Θ. Si $q(t)$ es la densidad arbitraria continua y positiva (dentro de Θ) de la distribución a priori, entonces*

$$\sqrt{n}(\hat{\theta}^* - \theta^*_Q) \underset{c.s.}{\to} 0,$$

donde θ^*_Q es la estimación bayesiana correspondiente a $q(t)$.

El carácter asintóticamente minimax de $\hat{\theta}^*$ puede ser establecido igualmente que en el teorema 25.7, con ayuda del análogo multidimensional del criterio asintóticamente minimax en el corolario 20.3:

$$\lim_{n \to \infty} \sup_{t \in \Gamma} \mathbf{M}_t n(\hat{\theta}^* - \theta) V(\hat{\theta}^* - \theta)^T = \sup_{t \in \Gamma} \sum I_{ij}^{-1}(\theta) v_{ij}$$

$$\|I_{ij}^{-1}(\theta)\| = I^{-1}(\theta),$$

y con ayuda del carácter uniforme de convergencia en (13), la cual se deducirá de los resultados del párrafo siguiente.

En el caso del parámetro multidimensional $\hat{\theta}^*$, cuando su dimensión k es grande, las propiedades de la optimalidad asintótica de θ deben utilizarse con cuidado. Es necesario observar que la relación n/k sea grande (el número de observaciones para un parámetro escalar). De lo contrario las deducciones pueden resultar erróneas.

Ejemplo 1. En el laboratorio se comprueba la concentración de n soluciones. Cada una de las n concentraciones desconocidas $\mu_1, \ldots, \mu_n$ se verifica dos veces. Se supone que la varianza σ^2 de todas n observaciones $(x_1, y_1) \ldots, (x_n, y_n)$ es igual, y que las propias observaciones son independientes y están distribuidas normalmente, así que

$$f_\theta(X) = \frac{1}{\sigma^{2n}(2\pi)^n} \exp\left\{-\frac{1}{2\sigma^2}\sum[(x_i - \mu_i)^2 + (y_i - \mu_i)^2]\right\},$$

donde

$$\theta = (\mu_1, \ldots, \mu_n, \sigma^2).$$

Las e.v.m. para μ_i aquí son iguales a

$$\hat{\mu}_i^* = \frac{1}{2}(x_i + y_i).$$

Es evidente que estas estimaciones no están desplazadas y no son conciliables. La e.v.m. para σ^2 es igual a

$$(\hat{\sigma}^2)^* = \frac{1}{4n}\sum (x_i - y_i)^2 \underset{P}{\to} \sigma^2/2 \text{ cuando } n \to \infty.$$

Esta estimación proporciona con gran fiabilidad un valor falso para el parámetro σ^2 (dos veces menor).

4. Cálculo aproximado de la e.v.m. El contenido de § 26 conservará por completo su validez en el caso multidimensional si por $[L''(X, t)]^{-1}$ entendemos la matriz inversa a $L''(X, t)$.

5. Propiedades de la e.v.m. al faltar las condiciones de regularidad (resultados de § 27). Las condiciones de conciliabilidad de θ, enunciadas en los teoremas 27.1—27.3, de hecho no están relacionadas con la dimensión. La demostración de estos teoremas se conserva por completo con una exactitud de hasta las modificaciones evidentes debidas al hecho de que el conjunto Θ ahora ha de ser recubierto (en virtud de la condición (A_c)) no por un número finito de intervalos, sino por un número finito de esferas. También se puede decir lo mismo en cuanto a los corolarios 27.1—27.4.

§ 29. Uniformidad respecto a θ, de las propiedades asintóticas de la relación de verosimilitud y de las estimaciones de verosimilitud máxima

En las investigaciones posteriores, principalmente en los §§ 13—15 de capítulo siguiente, serán útiles las afirmaciones de los §§ 24, 25 y 28 en su aspecto uniforme en cuanto a θ. La mayoría de estas afirmaciones (digamos, las que tratan de la $\mathbf{P}_\theta$-distribución límite de $(\hat{\theta}^* - \theta)\sqrt{n}$) han sido obtenidas suponiendo que θ es un punto registrado de Θ. Ahora nos interesará qué sucederá si θ no ha sido registrado y cambia junto con n. Está claro que en este caso junto con n también cambiarán las distribuciones $\mathbf{P}_\theta$, así que cada muestra X_n tendrá su "propia" distribución para $n = 1, 2, \ldots$

Llegamos, pues, al esquema de series (véase [11]), para el cual las enunciaciones de los principales teoremas del límite serán algo diferentes. En particular, la ley fuerte de los grandes números pierde, hablando en general, su sentido, ya que las variables aleatorias sujetas a investigación dejan de ser dadas (para diferentes n) en un espacio probabilístico.

1. **Ley uniforme de los grandes números y teorema central del límite.** Sea $X \in \mathbf{P}_\theta$, $\eta_{n,\theta} = \eta_n(X, \theta)$.

Definición 1. *Diremos que la sucesión* $\eta_{n,\theta}$ *converge uniformemente en probabilidad hacia la constante* $a(\theta)$, *si para cualquier* $\varepsilon > 0$, cuando $n \to \infty$,

$$\sup_{\theta \in \Theta} \mathbf{P}_\theta(|\eta_{n,\theta} - a(\theta)| > \varepsilon) \to 0.$$

Esta relación se escribirá en la forma "$\eta_{n,\theta} \underset{P_\theta}{\to} a(\theta)$ *uniformemente respecto a* θ".

Definición 2. Diremos que $\eta_{n,\theta}$ *converge en distribución hacia la variable aleatoria* η_θ *uniformemente respecto a* θ si para cualquier función continua y limitada φ, cuando $n \to \infty$,

$$\sup_\theta |\mathbf{M}_{\theta\varphi}(\eta_{n,\theta}) - \mathbf{M}_\varphi(\eta_\theta)| \to 0. \tag{1}$$

Esta relación es escribirá en la forma "$\eta_{n,\theta} \Rightarrow \eta_\theta$ *uniformemente respecto a* θ". Ese mismo sentido le conferiremos a la relación "$\eta_{n,\theta} \in \mathbf{G}_\theta$ uniformemente respecto a θ", donde $\mathbf{G}_\theta$ significa la distribución η_θ.

Le proponemos al lector que él mismo compruebe el hecho de que si las funciones de distribución η_θ son continuas uniformemente respecto a θ, la relación (1) es equivalente a

$$\sup_{\theta,\, x} |\mathbf{P}_\theta(\eta_{n,\theta} < x) - \mathbf{P}(\eta_\theta < x)| \to 0.$$

Nótese que la convergencia uniforme $\eta_{n,\theta} \underset{P_\theta}{\to} a(\theta)$ y la convergencia uniforme en distribución $\eta_{n,\theta} \Rightarrow a(\theta)$ hacia la variable aleatoria degenerada $a(\theta)$ son equivalentes.

Nótese también que para la convergencia uniforme conservarán su validez los principales teoremas de continuidad. Por ejemplo, si H es una función continua, de la convergencia uniforme $\eta_{n,\theta} \Rightarrow \eta_\theta$ se deduce la convergencia uniforme

$$H(\eta_{n,\theta}) \Rightarrow H(\eta_\theta). \tag{2}$$

Estas afirmaciones se deducen directamente de las definiciones.

En el Suplemento V hemos demostrado los siguientes teoremas "uniformes" del límite.

Supongamos que $X \in \mathbf{P}_\theta$ y que $a(x, \theta)$ es una función vectorial medible dada : $\mathscr{X} \times \Theta \to R^l$. Examinemos las sumas

$$s_n(\theta) = \sum a(x_i, \theta)$$

de los vectores aleatorios independientes que dependen del parámetro $\theta \in \Theta$

tanto directamente a través de la función $a(x, \theta)$, como también a través de la distribución de $x_i \in \mathbf{P}_\theta$.

Recordemos que la integral $\int \psi(x, \theta)\mathbf{P}_\theta(dx)$ se llama convergente uniformemente respecto a θ en la región Θ, si

$$\sup_{\theta\in\Theta} \int_{|\psi(x,\theta)|>N} |\psi(x, \theta)|\mathbf{P}_\theta(dx) \to 0$$

cuando $N \to \infty$.

Teorema 1 (ley uniforme de los grandes números). *Si la integral* $a(\theta) = \int a(x, \theta)\mathbf{P}_\theta(dx)$ *converge uniformemente respecto a* $\theta \in \Theta$, *entonces, cuando* $n \to \infty$,

$$\frac{s_n(\theta)}{n} \underset{P_\theta}{\to} a(\theta)$$

uniformemente respecto a θ.

Corolario 1. *Si la sucesión* $\{\theta_n\} \in \Theta$, *entonces en las condiciones del teorema* 1,

$$\mathbf{P}_{\theta_n}\left(\left|\frac{s_n(\theta_n)}{n} - a(\theta_n)\right| > \varepsilon\right) \to 0.$$

Este hecho será designado

$$\frac{s_n(\theta_n)}{n} - a(\theta_n) \underset{\mathbf{P}_{\theta_n}}{\to} 0.$$

Al examinar el teorema central del límite, para las sumas $s_n(\theta)$ será más cómodo suponer $a(\theta) = 0$. (Esto no es la limitación de la generalidad, ya que podemos examinar nuevos sumandos $a^1(x_i, \theta) = a(x_i, \theta) - a(\theta)$). Pongamos $\sigma^2(\theta) = \mathbf{M}_\theta(a^T(x_1, \theta)a(x_1, \theta))$ y designemos por $a_j(x_1, \theta)$, $j = 1, 2, \ldots, l$ las coordenadas de los vectores $a(x_1, \theta)$.

Teorema 2 (teorema central uniforme del límite). *Supongamos que las integrales* $\int(a_j^2(x, \theta))\mathbf{P}_\theta(dx)$, $j = 1, \ldots, l$ *convergen uniformemente en* Θ. *Entonces*

$$\eta_{n,\theta} = \frac{s_n(\theta)}{\sqrt{n}} \Rightarrow \eta_\theta \in \Phi_{0,\sigma^2(\theta)}$$

uniformemente respecto a θ.

2. Variantes uniformes de los teoremas de las propiedades asintóticas de la relación de verosimilitud y de las estimaciones de verosimilitud máxima. Nótese previamente que, al cumplirse las condiciones (RR), los resultados del § 23 serán uniformes respecto a θ por su propia forma, ya que los segundos miembros de las desigualdades en los teoremas 23.1 — 23.3 (y en los teoremas 28.1 — 28.3) no dependen de θ.

Pasemos a los resultados de los §§ 24 y 28 acerca del comportamiento asintótico de $Z(u/\sqrt{n})$.

Las afirmaciones de los lemas 24.1, 28.1, 24.2 y 28.2 pueden hacerse uniformes respecto a θ.

Lema 1. *Cuando* $\Delta \to 0$

$$\sup_{\theta} \mathbf{M}_\theta \omega_\Delta''(x_1) \to 0, \tag{3}$$

donde $\omega_\Delta(x_1)$ *es el módulo máximo de continuidad de las funciones* $l_{ij}''(x, \theta)$.

Demostración. La validez de (3) para un θ registrado ha sido demostrada en el lema 28.1. Si en este caso admitimos la ausencia de uniformidad respecto a θ, llegaremos al hecho de que existen $\varepsilon > 0$ y sucesiones $\theta_n \to \theta \in \Theta$, $\Delta_n \to 0$ tales, que

$$\mathbf{M}_{\theta_n} \omega_{\Delta_n}''(x_1) > \varepsilon. \tag{4}$$

Suponiendo, para abreviar, $\omega_{\Delta_n}''(x_1) = \omega''$, obtenemos

$$\mathbf{M}_{\theta_n}\omega'' = \mathbf{M}_{\theta_n}(\omega''; f_{\theta_n}(x_1) \leqslant 2f_\theta(x_1)) + \mathbf{M}_{\theta_n}(\omega''; f_{\theta_n}(x_1) > 2f_\theta(x_1), l(x_1) \leqslant N) + \mathbf{M}_{\theta_n}(\omega''; f_{\theta_n}(x_1) > 2f_\theta(x_1), l(x_1) > N).$$

Aquí el primer sumando no excede $2\mathbf{M}_\theta\omega''$ y converge a cero en virtud del lema 28.1. El segundo sumando no supera $2NJ_n$, donde

$$J_n = \int_{f_{\theta_n}(x) > 2f_\theta(x)} f_{\theta_n}(x)\mu(dx) = 1 - \int_{f_{\theta_n}(x) \leqslant 2f_\theta(x)} f_{\theta_n}(x)\mu(dx) \to 0$$

según el teorema de la convergencia mayorada. Por fin, el último sumando no supera $\mathbf{M}_{\theta_n}(2l(x_1); l(x_1) > N)$ y, en virtud de (RR), puede hacerse, escogiendo N, tan pequeño como se quiera. Hemos obtenido la contradicción con (4), lo cual demuestra el lema.

Lema 2. *La afirmación del lema* 28.2 *se conservará si la convergencia casi segura en ella se sustituye por la convergencia* $\gamma_n(\delta_n, \theta) \underset{\mathbf{P}_\theta}{\to} 0$, $\gamma_n(\delta_n, \hat{\theta}^*) \underset{\mathbf{P}_\theta}{\to} 0$ uniforme respecto a θ.

Demostración. Seguiremos la demostración del lema 28.2. Señalemos previamente que, en virtud del teorema 1 y de la convergencia uniforme de la integral en (RR),

$$L''(X, \theta)/n \underset{\mathbf{P}_\theta}{\to} -I(\theta)$$

uniformemente respecto a θ (la convergencia de las matrices se entiende por elementos). Además, de los teoremas 23.3 y 28.3 se deduce que $\hat{\theta}^* \underset{P_\theta}{\to} \theta$ uniformemente respecto a θ. De aquí se desprende que en la relación $\gamma_n(\delta_n, \theta) \underset{P_\theta}{\to} 0$ (véase el lema 28.2) podemos sustituir $I(\theta)$ por $L''(\theta)/n$ y por $I(\hat{\theta}^*)$.

En virtud de la desigualdad (28.7), el problema de estimación de $\gamma_n(\delta_n, \theta)$ se reduce a la estimación de

$$\bar{\omega}''_{\delta_n}(X) = \frac{1}{n}\sum_{i=1}^{n} \omega''_{\delta_n}(x_i, \theta),$$

donde $\omega''_{\Delta}(x, \theta)$ es el módulo máximo de continuidad de las funciones $l''_{ij}(x, \theta)$. De la desigualdad de Chébishev obtenemos

$$\sup_{\theta} \mathbf{P}_\theta(\bar{\omega}''_{\delta_n}(X) > \varepsilon) \leqslant \frac{1}{\varepsilon} \sup_{\theta} \mathbf{M}_\theta\, \omega''_{\delta_n}(x_1, \theta).$$

Pero en virtud del lema 1, $\sup\limits_{\theta} \mathbf{M}_\theta \omega''_{\Delta}(x_1, \theta) \to 0$ cuando $\Delta \to 0$. Esto demuestra que

$$\bar{\omega}''_{\delta_n}(X) \underset{P_\theta}{\to} 0, \quad \gamma_n(\delta_n, \theta) \underset{P_\theta}{\to} 0 \tag{5}$$

uniformemente respecto a θ.

Luego, de las desigualdades (24.6) resulta que el problema de estimación de $\gamma_n(\delta_n, \hat{\theta}^*)$ se reduce a la estimación de $\bar{\omega}''_{\delta_n+|\theta^*-\theta|}(X)$. Como $\hat{\theta}^* - \theta \underset{P_\theta}{\to} 0$ uniformemente respecto a θ, de (5) obtenemos que

$$\bar{\omega}''_{\delta_n+|\theta^*-\theta|}(X) \underset{P_\theta}{\to} 0, \quad \gamma_n(\delta_n, \hat{\theta}^*) \underset{P_\theta}{\to} 0$$

uniformemente respecto a θ. $\lhd$

Teorema 3 (análogo del teorema 28.4). *Al cumplirse las condiciones (RR), las afirmaciones del teorema 28.4 se conservarán en las modificaciones siguientes:* $\varepsilon_n(X, \theta) \underset{P_\theta}{\to} 0$ *uniformemente respecto a* θ, $\xi_n \in \Phi_{0,I(\theta)}$, $2Y(u^*) \in H_k$ *uniformemente respecto a* θ.

La demostración del teorema se basa por completo en el lema 2, así como la demostración del teorema 28.4 se basa en el lema 28.2. Por eso la demostración requerida se obtiene mediante la introducción de modificaciones evidentes en la demostración del teorema 28.4, relacionadas con la sustitución (que resulta del lema 28.2) de la convergencia $\varepsilon_n(X, \theta) \underset{\text{c.s.}}{\to} 0$ por la convergencia uniforme $\varepsilon_n(X, \theta) \underset{P_\theta}{\to} 0$. Además, hay que añadir que

$$\xi_n = \frac{1}{n}\sum_{i=1}^{n} l(x_i, \theta) \in \Phi_{0,I(\theta)}$$

uniformemente respecto a θ, en virtud del teorema 2 y de la convergencia uniforme (28.6) de la integral $I(\theta)$ (ésta es la matriz de segundos momentos

para $l'(x_1, \theta)$), la cual se desprende de las condiciones (*RR*) (véase el Suplemento VI). De aquí y de las observaciones referentes a (2) obtenemos la convergencia uniforme

$$2Y(u^*) \Subset \mathbf{H}_k. \ \lhd$$

Las mismas modificaciones que en el teorema 3 (en comparación con el teorema 28.4) pueden ser introducidas en los teoremas 28.5 y 28.6.

Citemos aquí los dos siguientes corolarios del teorema 3.

Teorema 4.

$$u^* = \sqrt{n}(\hat{\theta}^* - \theta) \Subset \Phi_{0,I^{-1}(\theta)} \tag{6}$$

uniformemente respecto a θ. *En este caso, para cualquier función* $w(x)$ *continua casi por doquier respecto a la medida de Lebesgue y tal que* $|w(x)| < Ce^{\beta|x|^2/2}$ *(el valor de* $\beta > 0$ *ha sido definido en el teorema* 28.2), *se cumple*

$$\sup_\theta |\mathbf{M}_\theta w(u^*) - \mathbf{M}w(\eta_\theta)| \to 0, \tag{7}$$

donde $\eta_\theta \Subset \Phi_{0,I^{-1}(\theta)}$.

Demostración. La primera afirmación se deduce de las relaciones

$$u^* = \xi_n I^{-1}(\theta)(E + \varepsilon_n(X, \theta)),$$
$$|\varepsilon_n(X, \theta)| \underset{P_\theta}{\to} 0, \quad \xi_n \Subset \Phi_{0,I(\theta)},$$

uniformes respecto a θ y contenidas en el teorema 3.

Para demostrar la segunda afirmación admitamos que (7) no es cierta. Entonces habrá $\delta > 0$ y sucesiones $\theta_n \to \theta \in \Theta$ tales, que

$$|\mathbf{M}_{\theta_n} w(u^*) - \mathbf{M}w(\eta_{\theta_n})| > \delta \tag{8}$$

para todos n.

Pero $\Phi_{0,I^{-1}(\theta_n)} \Rightarrow \Phi_{0,I^{-1}(\theta)}$ y, por consiguiente, en virtud de (6), la $\mathbf{P}_{\theta_n}$-distribución $u^*(w(u^*))$ converge débilmente a la distribución $\eta_\theta(w(\eta_\theta))$. Además, según el corolario 23.2 (véase también el § 28),

$$\sup_\theta \mathbf{M}_\theta w^{3/2}(u^*) \leqslant \sup_\theta \mathbf{M}_\theta \exp\{3(u^*)^2\beta/4\} < c_1 < \infty.$$

De aquí y de los teoremas de continuidad para los momentos se deduce que

$$\mathbf{M}_{\theta_n} w(u^*) \to \mathbf{M}w(\eta_\theta).$$

En vista de que $\mathbf{M}w(\eta_{\theta_n}) \to \mathbf{M}w(\eta_\theta)$, la relación obtenida contradice (8). $\lhd$

Sea $A_n \subset \mathscr{X}^n$.

Teorema 5. *Si* $\mathbf{P}_\theta(A_n) \to 0$, *entonces para cualquier* N *registrado,*

$$\sup_{|u| \leq N} \mathbf{P}_{\theta + u/\sqrt{n}}(A_n) \to 0.$$

Esta propiedad de las sucesiones de las distribuciones $\mathbf{P}_{\theta + u/\sqrt{n}}$ cuando $n \to \infty$ se llama *contigualidad* (véase [81]). La utilizaremos en el capítulo 3.

Demostración. Tenemos

$$\mathbf{P}_{\theta + u/\sqrt{n}}(A_n) = \mathbf{M}_\theta \{Z(u/\sqrt{n});\ A_n\} \leq$$
$$\leq \mathbf{M}_\theta(Z(u/\sqrt{n});\ A_n \cap \{Y(u) \leq c\}) + \mathbf{P}_{\theta + u/\sqrt{n}}(Y(u) > c) \leq$$
$$\leq e^c \mathbf{P}_\theta(A_n) + \mathbf{P}_{\theta + u/\sqrt{n}}(Y(u) > c).$$

Como $\mathbf{P}_\theta(A_n) \to 0$, para demostrar el teorema debemos examinar sólo $\sup\limits_{|u| \leq N} \mathbf{P}_{\theta + u/\sqrt{n}}(Y(u) > c)$. Según el teorema 3,

$$Y(u) = (\xi_n,\ u) - \frac{1}{2} uI(\theta)u^T(1 + \varepsilon_n(X,\ \theta + u/\sqrt{n})) \Subset \Phi_{-\frac{1}{2}\sigma^2, \sigma^2} \tag{9}$$

uniformemente respecto a u, donde $\sigma^2 = uI(\theta)u^T \leq N^2 \Lambda_k(\theta)$ cuando $|u| \leq N$, y $\Lambda_k(\theta)$ es el número máximo propio de la matriz $I(\theta)$. Como $\Phi_{-\frac{1}{2}\sigma^2, \sigma^2}((c,\ \infty)) \leq \Phi_{0, \sigma^2}((c,\ \infty))$, entonces, en virtud de la uniformidad en (9),

$$\lim_{n \to \infty} \sup_{|u| \leq N} \mathbf{P}_{\theta + u/\sqrt{n}}(Y(u) > c) \leq \sup_{|u| \leq N} \Phi_{0, \sigma^2}((c,\ \infty)) = \Phi_{0, N^2 \Lambda_k(\theta)}((c,\ \infty)).$$

Eligiendo c, este valor puede hacerse tan pequeño como se quiera. ◁

3. Algunos corolarios.

1) En el § 25 hemos enunciado el teorema 25.3 en el que se afirma, en particular, que $\hat{\theta}^* \in \bar{K}^\circ$, donde $\bar{K}^\circ$ es la clase de estimaciones asintóticamente centrales, la cual es definida por la relación (se examina el caso unidimensional)

$$\mathbf{P}_\theta(\hat{\theta}^* > \theta) \to 1/2$$

uniformemente respecto a θ. Del teorema 4 se deduce que la parte mencionada del teorema 25.3 es cierta, así que

$$\mathbf{P}_\theta(\hat{\theta}^* > \theta) = \mathbf{P}_\theta(\sqrt{n}(\hat{\theta}^* - \theta)I^{-1/2}(\theta) > 0) \to \Phi_{0,1}\ ((0,\ \infty)) = 1/2$$

uniformemente respecto a θ. ◁

2) En el § 25 hemos enunciado el teorema 25.7 acerca del carácter asintóticamente minimax de $\hat{\theta}^*$. Para demostrar este teorema sólo queda establecer la validez del lema 25.1 de que

$$\lim_{n \to \infty} \sup_{\theta \in \Gamma} \mathbf{M}_\theta n(\hat{\theta}^* - \theta)^2 = \sup_{\theta \in \Gamma} I^{-1}(\theta), \tag{10}$$

donde Γ es cualquier segmento de Θ. Pero esta afirmación es el corolario directo de la convergencia de $\mathbf{M}_\theta n(\hat{\theta}^* - \theta)^2 \to I^{-1}(\theta)$, uniforme respecto a $\theta \in \Theta$, la cual hace válido el paso límite bajo el signo $\sup\limits_{\theta\in\Gamma}$:

$$\lim_{n\to\infty} \sup_{\theta\in\Gamma} \mathbf{M}_\theta n(\hat{\theta}^* - \theta)^2 = \sup_{\theta\in\Gamma} \lim_{n\to\infty} \mathbf{M}_\theta n(\hat{\theta}^* - \theta)^2 = \sup_{\theta\in\Gamma} I^{-1}(\theta). \quad \triangleleft$$

La afirmación, que es análoga a (10) y asegura el carácter asintóticamente minimax de $\hat{\theta}^*$, tendrá lugar, evidentemente, también en el caso multidimensional:

$$\lim_{n\to\infty} \sup_{\theta\in\Gamma} \mathbf{M}_\theta n(\hat{\theta}^* - \theta) V(\hat{\theta}^* - \theta)^T = \sup_{\theta\in\Gamma} \sum v_{ij} I_{ij}^{-1}(\theta),$$

$$\| I_{ij}^{-1}(\theta) \| = I^{-1}(\theta)$$

para cualquier matriz V.

§ 30*. Acerca de los problemas estadísticos relacionados con las muestras de volumen aleatorio. Estimación sucesiva

El hecho de que las muestras de volumen aleatorio surgen en la práctica y son naturales, es confirmado por el ejemplo 18.3. Otro ejemplo está relacionado con la llamada estimación sucesiva (o progresiva), que se emplea en los casos cuando podemos realizar observaciones sucesivas, es decir, una tras otra, y cuando estamos interesados en minimizar el número de tales observaciones, digamos, debido a su alto precio. En este caso, además de la regla de estimación (o sea, de construcción de la estimación θ^*) debemos establecer la regla de interrupción del experimento. Estas reglas pueden ser diferentes: por ejemplo, podemos sumar los precios dados $c(\mathrm{x}_i)$ de las observaciones x_i hasta agotarse cierta cantidad admisible t. En este caso el momento ν de interrupción (número de la última observación o volumen de la muestra) será determinado como

$$\nu = \min \left\{ k: \sum_{i=1}^{k} c(\mathrm{x}_i) \geqslant t \right\},$$

esto es "el tiempo del primer del nivel t" en errar con saltos $c(\mathrm{x}_i)$ (véase [11], capítulo 8). Se pueden sumar las "informaciones" $I(\mathrm{x}_i, \theta) = (l'(\mathrm{x}_i, \theta))^2$ e interrumpir las observaciones cuando sea alcanzado otra vez cierto nivel dado, etc.

En estos ejemplos ν es un momento markoviano, o sea, $\{\nu > n\} \in \sigma(\mathrm{x}_1, \ldots, x_n)$, que constituye una de las suposiciones principales al examinar los problemas de estimación sucesiva. Al hacer tal suposición y al cumplirse varias condiciones adicionales menos esenciales, la desigualdad

de Rao — Cramer será conservada en la forma siguiente:

$$\mathbf{D}_\theta \theta^* \geqslant \frac{1}{I(\theta)\mathbf{M}\nu},$$

donde $\theta^* = \theta^*(x_1, \ldots, x_\nu)$ es la estimación no desplazada de θ, $I(\theta)$, es decir, la información de Fisher. La demostración de esta desigualdad es análoga a las demostraciones del § 16, para calcular la información de Fisher, contenida en la muestra $(x_1, \ldots, x_\nu)$, sólo se necesita utilizar la identidad de Wald (véase [11]).

Si ν depende de cierto parámetro t, como ocurrió en el ejemplo 18.3, así que $\nu \to \infty$ casi siempre cuando $t \to \infty$, entonces es posible construir las estimaciones asintóticamente óptimas con un error estándar asintóticamente equivalente a $(I(\theta)\mathbf{M}\nu)^{-1}$.

§ 31. Estimación por intervalo

1. Definiciones. Hasta ahora hemos estudiado las propiedades y los métodos de búsqueda de las mejores estimaciones *puntuales* de un parámetro desconocido que determina la distribución $\mathbf{P}_\theta$ de la familia $\mathscr{P} = \{\mathbf{P}_\theta\}$, correspondiente a la muestra X. Las estimaciones puntuales se utilizan en los casos cuando debemos llamar cierto número θ^* destinado al uso en vez de θ desconocido.

No obstante, también tiene gran aplicación otro enfoque de la cuestión.

Consideraremos θ como parámetro escalar (el caso multidimensional será examinado en el punto 6). Como sabemos, no es posible determinar exactamente θ basándose en una muestra dada. Pero podríamos tratar de indicar tal intervalo (θ^-, θ^+), el cual, con una probabilidad dada bastante alta, sea capaz de recubrir el valor desconocido de θ. En este caso es indudable que cuanto más estrecho sea este intervalo tanto mejor será. En muchos problemas se exige de antemano, digamos, aumentando el volumen de la muestra, construir tal intervalo (θ^-, θ^+) cuya anchura no supere las dimensiones dadas.

Definición 1. Supongamos que para $\varepsilon > 0$ dado existen variables aleatorias $\theta^\pm = \theta^\pm(\varepsilon, X)$ tales que

$$\mathbf{P}_\theta(\theta^-(\varepsilon, X) < \theta, \quad \theta^+(\varepsilon, X) > \theta) \geqslant 1 - \varepsilon. \tag{1}$$

Entonces el intervalo (θ^-, θ^+) se llama *intervalo confidencial para* θ *de nivel* $1 - \varepsilon$.

Es evidente que (1) se puede escribir en la forma

$$\mathbf{P}_\theta(\theta^- < \theta < \theta^+) \geqslant 1 - \varepsilon.$$

El suceso que aquí está bajo el signo de probabilidad, consiste en que el *intervalo aleatorio* (θ^-, θ^+) *ha cubierto el valor desconocido de* θ. Leer

este suceso como "θ toma un valor perteneciente al intervalo (θ^-, θ^+)" sería un poco menos exacto, ya que θ, hablando en general, no es aleatorio.

Los valores de $\theta^\pm$ se denominan *fronteras de los intervalos confidenciales*, y el número $1 - \varepsilon$, *coeficiente o nivel de confianza*.

Por lo tanto, la diferencia entre la estimación por intervalo y la estimación puntual consiste en lo siguiente.

1) El intervalo confidencial como estimación es "menos exacto", ya que se señala un conjunto entero de posibles valores de θ.

2) Por otro lado, la afirmación "$\theta \in (\theta^-, \theta^+)$ con probabilidad $\geqslant 1 - \varepsilon$" es real, mientras que el suceso $\theta = \theta^*$ tiene, por lo general, una probabilidad igual a cero.

En calidad de ε suele escogerse un número pequeño. Basándose en éste, se construyen $\theta^\pm(\varepsilon, X)$ y luego, basándose en la muestra, se declara que $\theta \in (\theta^-(\varepsilon, X), \theta^+(\varepsilon, X))$. Procediendo de este modo nos equivocaremos en una larga serie de experimentos, aproximadamente en el $100\,\varepsilon\%$ de todos los casos. Por ejemplo, si $\varepsilon = 0{,}001$, el error puede ocurrir una vez en 1000 casos, aproximadamente.

Declarando justa la relación $\theta \in (\theta^-, \theta^+)$, utilizamos el hecho de que si cierto suceso tiene la probabilidad ε y este ε es pequeño, entonces prácticamente es imposible que tal suceso se produzca durante un solo experimento. Un pasajero, tomando el avión cree intuitivamente en ello con seguridad. Le basta saber que la probabilidad de que el vuelo se termine felizmente es bastante alta (a pesar de que conoce que esta probabilidad no es igual a 1). Precisamente tal enfoque es la base para construir muchos procedimientos estadísticos.

Destaquemos primeramente un caso, cuando la construcción de los intervalos confidenciales es sobre todo natural y puede ser realizada sin grandes dificultades. Es el llamado caso bayesiano que ya hemos examinado en los §§ 10, 11 y 20.

2. Construcción de intervalos confidenciales en el caso bayesiano. Aquí supondremos que el parámetro θ se escoge *aleatoriamente*, con una densidad a priori conocida de distribución $q(t)$ respecto a cierta medida λ en Θ. Luego se realiza la muestra $X \in \mathbf{P}_\theta$ y necesitamos construir el intervalo confidencial para el valor elegido de θ.

Si se cumple la condición (A_μ), en este caso, como sabemos del § 10, existe una distribución a posteriori de θ (convencional respecto a X) que tiene una densidad de

$$q(t/X) = \frac{f_t(X)q(t)}{\int f_u(X)q(u)\lambda(du)}$$

respecto a la medida λ. Esto quiere decir que en calidad de $\theta^\pm(\varepsilon, X)$ es

suficiente tomar dos números cualesquiera $\theta^{\pm}$, para los cuales

$$\int_{\theta^-}^{\theta^+} q(u/X)\lambda(du) = 1 - \varepsilon$$

(o bien $\geqslant 1 - \varepsilon$ si $\int_{-\infty}^{t} q(u/X)\lambda(du)$ cambia al variar t discretamente). En otros términos, en calidad de θ^- y θ^+ conviene tomar las cuantilas de distribución a posteriori que tienen los órdenes $1 - \varepsilon_2$ y ε_1, respectivamente, para todos ε_1 y ε_2, tales que $\varepsilon_1 + \varepsilon_2 = \varepsilon$.

Aquí, a distinción del caso no bayesiano, en la relación $\theta^- \leqslant \theta \leqslant \theta^+$ son aleatorios todos los tres elementos : las fronteras del intervalo de $\theta^{\pm}$ y la propia magnitud θ.

No es difícil ver que en el procedimiento descrito existe cierta arbitrariedad relacionada con la elección de los números ε_1 y ε_2. A veces esta arbitrariedad es eliminada por el propio planteamiento del problema, por ejemplo, cuando nos es importante establecer únicamente la frontera confidencial superior o inferior. En este caso conviene poner igual a 0 uno de los números ε_1, ε_2 y hacer infinita la frontera respectiva. Sin embargo, si las fronteras desempeñan un papel simétrico, es natural escoger ε_i de modo que el intervalo (θ^-, θ^+) se haga más corto en la medida de lo posible. Para las distribuciones $q(t/X)$ próximas a las distribuciones simétricas, esto se alcanza cuando $\varepsilon_1 = \varepsilon_2 = \varepsilon/2$.

3. Construcción de intervalos confidenciales en el caso general. Intervalos confidenciales asintóticos. Los principales métodos de construcción de intervalos confidenciales se basan en la utilización de estimaciones puntuales. Examinemos al principio el enfoque asintótico de la construcción de intervalos confidenciales.

Definición 2. Supongamos que $X = [X_\infty]_n \in \mathbf{P}_\theta$ y que para $\varepsilon > 0$ establecido existen variables aleatorias $\theta^{\pm}(\varepsilon, X)$ tales que

$$\liminf_{n\to\infty} \mathbf{P}_\theta(\theta^-(\varepsilon, X) < \theta < \theta^+(\varepsilon, X)) \geqslant 1 - \varepsilon. \tag{2}$$

En este caso el intervalo (θ^-, θ^+) se llama intervalo *asintótico confidencial de nivel* $1 - \varepsilon$.

En esta definición es necesario subrayar que en realidad se trata de la sucesión de intervalos (θ_n^-, θ_n^+) determinados para cada n. Formalmente, el concepto de intervalo asintótico confidencial, con arreglo a una muestra de volumen registrado, es insustancial. No obstante, la relación (2) se utiliza con grandes n al igual que se utiliza el teorema central del límite para el cálculo aproximado de las distribuciones de las sumas de un número finito de variables aleatorias.

En los apartados precedentes hemos visto que la mayoría de las estimaciones puntuales examinadas eran asintóticamente normales. Más abajo se expone la construcción de los intervalos asintóticos confidenciales basados en tales estimaciones.

Sea θ^* la estimación asintóticamente normal:

$$(\theta^* - \theta)\sqrt{n} \in \Phi_{0,\sigma^2(\theta)}, \tag{3}$$

y $\sigma(\theta)$ es una función continua. Como $\theta^* \to \theta$, la última condición significa que $\sigma(\theta^*) \underset{P}{\to} \sigma(\theta)$. De aquí y de (3), según el segundo teorema de continuidad, resulta que

$$\frac{(\theta^* - \theta)\sqrt{n}}{\sigma(\theta^*)} \in \Phi_{0,1}. \tag{4}$$

Designemos por λ_δ la cuantila de distribución normal de orden $1 - \delta$, o sea, un número tal que $\Phi_{0,1}((-\infty, \lambda_\delta)) = 1 - \delta$, o bien $\mathbf{P}(|\xi| < \lambda_\delta) = 1 - 2\delta$ si $\xi \in \Phi_{0,1}$. Al disponer de $\varepsilon > 0$ registrado, para $\lambda_{\varepsilon/2}$ introduzcamos temporalmente una designación más breve, suponiendo

$$\lambda_{\varepsilon/2} = \beta.$$

Entonces de (4) se deduce

$$\lim_{n\to\infty} \mathbf{P}_\theta \left(\left| \frac{(\theta^* - \theta)\sqrt{n}}{\sigma(\theta^*)} \right| < \beta \right) = 1 - \varepsilon.$$

Pero esta relación se puede escribir en la forma

$$\lim_{n\to\infty} \mathbf{P}_\theta(\theta^* - \beta\sigma(\theta^*)/\sqrt{n} < \theta < \theta^* + \beta\sigma(\theta^*)/\sqrt{n}) = 1 - \varepsilon.$$

Ahora bien, los números

$$\theta^\pm = \theta^* \pm \beta\sigma(\theta^*)/\sqrt{n} \tag{5}$$

satisfacen la definición 2 y, por consiguiente, son las fronteras del intervalo asintótico confidencial de nivel $1 - \varepsilon$.

Si ahora, para la muestra X dada y registrada, de volumen n, construimos el intervalo (5), su nivel real se distinguirá, hablando en general, de ε, pero se distinguirá poco si n es bastante grande. Por eso los intervalos asintóticos confidenciales deben tratarse con cierto cuidado, aclarando previamente a partir de qué n la probabilidad del suceso $\{\theta \in (\theta^-, \theta^+)\}$ es con bastante exactitud aproximada por el valor límite. Por regla general, cuanto menor sea ε tanto mayor será la exigencia en cuanto al volumen de la muestra n. El volumen necesario también depende de la distribución $\mathbf{P}_\theta$ y de la estadística θ^*.

Ejemplo 1. Supongamos que $X \in \Gamma_{\alpha,1}$ y que utilizamos la estimación eficiente $\alpha^* = \frac{n-1}{n\bar{x}}$. En los ejemplos 4.1 y 16.1 hemos establecido que

$$\mathbf{M}_\alpha\alpha^* = \alpha, \quad \mathbf{D}_\alpha\alpha^* = \frac{\alpha^2}{n-2},$$

así que aquí $\sigma^2(\alpha) = \alpha^2$. La relación (5) nos da

$$\alpha^\pm = \frac{n-1}{n\bar{x}}(1 \pm \beta/\sqrt{n}). \tag{6}$$

¿A qué realmente es igual el nivel de este intervalo?

Necesitamos hallar $\Gamma_{\alpha,1}$, o sea, la probabilidad de la desigualdad

$$\frac{n-1}{n\bar{x}}(1 - \beta/\sqrt{n}) < \alpha < \frac{n-1}{n\bar{x}}(1 + \beta/\sqrt{n})$$

o bien, que es lo mismo, la probabilidad de la desigualdad

$$1 - \beta/\sqrt{n} < \frac{n\alpha\bar{x}}{n-1} < 1 + \beta/\sqrt{n},$$

donde $n\alpha\bar{x} \in \Gamma_{1,n}$. Como α es el parámetro de escala, entonces $2n\alpha\bar{x} \in \Gamma_{1/2,n} = \mathbf{H}_{2n}$. Así pues, el nivel exacto del intervalo (6) es igual a

$$\int_{2(n-1)(1-\beta/\sqrt{n})}^{2(n-1)(1+\beta/\sqrt{n})} \gamma_{1/2,n}(x)dx, \tag{7}$$

donde $\gamma_{1/2,n}$ está definido en el § 2*).

Cuando $\varepsilon = 0{,}05$ y $n = 30$, tenemos $\beta = 1{,}96$, $(n-1)(1 - \beta/\sqrt{n})/n \approx \approx 0{,}6201$, $(n-1)(1 + \beta/\sqrt{n})/n \approx 1{,}3126$.

Ahora bien, el intervalo asintótico confidencial de nivel $1 - \varepsilon = 0{,}95$ con arreglo al caso $n = 30$, es el intervalo $(0{,}620/\bar{x},\ 1{,}313/\bar{x})$.

Si hacemos uso de las tablas de distribución χ^2 con 60 grados de libertad, en virtud de (7) descubriremos que el nivel exacto de significación de este intervalo confidencial constituye (con una exactitud de hasta tres signos) $0{,}937 = 1 - 0{,}063$. En este caso los "aportes" de los extremos izquierdo y derecho del referido intervalo no son equivalentes ni mucho menos (compárese con la aproximación normal) y constituyen 0,010 y 0,053, respectivamente.

Para $n = 50$ el intervalo asintótico confidencial de nivel, igual a 0,95, tendrá la forma $(0{,}708/\bar{x},\ 1{,}252/\bar{x})$. El nivel real de su significación será

*) La observación de que $\Gamma_{1/2,n} = \mathbf{H}_{2n}$ es útil, ya que permite, para el cálculo de $\Gamma_{\alpha,\lambda}$ (si 2λ es entero), utilizar las tablas de la distribución χ^2 dadas en el suplemento, así como en muchos otros manuales de estadística matemática.

igual a 0,942 = 1 − 0,058 (los aportes equivalen a 0,014 y 0,044, respectivamente). Está claro que si continuamos aumentando n, dichos aportes se aproximarán con 0,025.

Volvamos a examinar el intervalo confidencial (5) que hemos construido con ayuda de la estimación asintóticamente normal θ^*. A distinción del caso bayesiano, aquí hay una arbitrariedad relacionada con la elección de la estimación θ^*. La forma de las fronteras del intervalo muestra que se pueden obtener las dimensiones dadas del intervalo, tanto aumentando el volumen de la muestra n (lo que por diferentes causas no siempre es realizable) como disminuyendo posiblemente $\sigma(\theta^*)$. Aquí llegamos a la conclusión importante de que siendo iguales los volúmenes de la muestra, la estimación de menor dispersión $\sigma(\theta)$ dará el mejor intervalo confidencial. Ahora bien, *los mejores intervalos asintóticos confidenciales se obtendrán al utilizar las estimaciones asintóticamente eficientes.*

Siempre que se cumplan las condiciones (RR) y que θ^* pertenezca a la clase $\tilde{K}_0 \cap K_{\Phi,2}$ (véanse los §§ 8 y 16) el mejor intervalo asintótico confidencial tendrá las siguientes fronteras:

$$\theta^{\pm} = \theta^* \pm \beta/\sqrt{nI(\theta^*)},$$

donde θ^* es cualquier estimación asintóticamente eficiente, por ejemplo, la e.v.m.

Algunos otros métodos de construcción de intervalos asintóticos confidenciales se examinarán en el punto 6.

4. Construcción del intervalo confidencial exacto mediante una estadística dada. Supongamos que en calidad de estadística hemos escogido la estimación θ^*. Entonces, mediante esta estimación, sería natural buscar el intervalo confidencial simétrico de nivel $1 - \varepsilon$ en la forma $\theta^* \pm \Delta(\varepsilon, X)$ o en la forma $\theta^*(1 \pm \Delta(\varepsilon, X))$, así como se hizo en el ejemplo antes examinado. No obstante, si tratamos de realizar este plan, resultará que la cosa no es tan simple, ya que en el caso general las fronteras $\pm\Delta(\varepsilon,, X)$ dependerán del parámetro desconocido θ: pues $\Delta(\varepsilon, X)$ debe ser elegido de la condición

$$\mathbf{P}_\theta(\theta^* - \Delta(\varepsilon, X) < \theta < \theta^* + \Delta(\varepsilon, X)) \geqslant 1 - \varepsilon,$$

donde θ aquí entra, de manera esencial y muy compleja, antes que nada a través de la propia distribución $\mathbf{P}_\theta$.

Por eso, para construir los intervalos confidenciales mediante una estimación dada θ^*, se necesita cierta estructura especial.

En la construcción expuesta más abajo, a la par con la estimación θ^* puede participar cualquier estadística S. Designemos con el símbolo $\mathbf{G}_\theta$ la distribución de S y pongamos $G_\theta(x) = \mathbf{G}_\theta((-\infty, x))$.

Definición 3. Diremos que la estadística S, *en cuanto a su distribución, depende monótonamente de* θ si para todos x, $\theta_1 < \theta_2$

$$\mathbf{G}_{\theta_1}((x, \infty)) \leqslant \mathbf{G}_{\theta_2}((x, \infty))$$

o bien, que es lo mismo,

$$G_{\theta_1}(x) \geqslant G_{\theta_2}(x). \tag{8}$$

Todas las estimaciones razonables θ^* suelen poseer esta propiedad.

Si la dependencia monótona $G_\theta(x)$ de θ es continua, entonces la ecuación

$$G_\theta(x) = \gamma$$

es siempre resoluble respecto a θ para cada $\gamma \in (0, 1)$. Designemos por $b(x, \gamma)$ la solución de esta ecuación.

Teorema 1. *Si* $\varepsilon_1 + \varepsilon_2 = \varepsilon$, *la estadística* S, *en cuanto a su distribución, depende monótonamente de* θ, *y la función* $G_\theta(x)$ *es continua respecto a* θ *y* x, *entonces los valores*

$$\theta^- = b(S, 1 - \varepsilon_2), \quad \theta^+ = b(S, \varepsilon_1)$$

formarán el intervalo confidencial de nivel $1 - \varepsilon$.

La demostración del teorema es casi evidente. Utilicemos el hecho de que si la función de distribución $F(x)$ es continua y $\xi \in F$, entonces $F(\xi) \in \mathbf{U}_{0,1}$ $(\mathbf{P}(F(\xi) < x) = \mathbf{P}(\xi < F^{-1}(x)) = F(F^{-1}(x)) = x)$. En virtud de esta observación, $G_\theta(S) \in \mathbf{U}_{0,1}$ y, por lo tanto,

$$\mathbf{P}_\theta(\varepsilon_1 < G_\theta(S) < 1 - \varepsilon_2) = 1 - \varepsilon,$$
$$\mathbf{P}_\theta(b(S, 1 - \varepsilon_2) < \theta < b(S, \varepsilon_1)) = 1 - \varepsilon. \triangleleft$$

Con frecuencia es cómodo realizar en dos etapas la "inversión" de la función $G_\theta(S)$, utilizada en el teorema. Primeramente $G_\theta(x)$ se invierte respecto a x, o sea, se determinan las cuantilas $G_\theta^{-1}(\gamma)$ como soluciones de las ecuaciones $G_\theta(x) = \gamma$, y luego se resuelven, respecto a θ, las ecuaciones

$$G_\theta^{-1}(\varepsilon_1) = S, \quad G_\theta^{-1}(1 - \varepsilon_2) = S.$$

Tales soluciones siempre existirán, ya que, según los datos del teorema, $G_\theta^{-1}(\gamma)$ depende monótona y continuamente de θ.

En la fig. 3 se muestran las curvas $y = G_\theta^{-1}(\varepsilon_1)$ e $y = G_\theta^{-1}(1 - \varepsilon_2)$ que definen para cada θ el campo de valores y, cuya probabilidad de entrar en el mismo, para cierta estimación $S = \theta^*$, es igual a $1 - \varepsilon$. Como ya hemos señalado, el procedimiento de construcción del intervalo confidencial es la inversión de las funciones

$$y = G_\theta^{-1}(\varepsilon_1), \quad y = G_\theta^{-1}(1 - \varepsilon_1),$$

o sea, la determinación de los puntos de intersección de las curvas de nivel $y = S$ que les corresponden. Los puntos de intersección obtenidos dan precisamente el intervalo requerido (θ^-, θ^+).

Si la condición de continuidad de $G_\theta(x)$ no se cumple, lo cual tendrá lugar para variables aleatorias discretas S, entonces, en general, el procedimiento expuesto y la afirmación del teorema 1 conservarán su validez, con la única diferencia de que, al definir respectivamente las cuantilas $G_\theta^{-1}(\gamma)$,

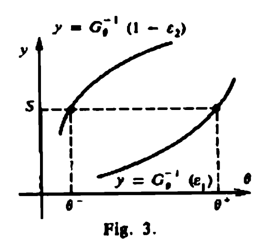


Fig. 3.

es necesario satisfacer la desigualdad $G_\theta((G_\theta^{-1}(\varepsilon_1), G_\theta^{-1}(1 - \varepsilon_2))) \geqslant 1 - \varepsilon$ en vez de la cual antes hemos tenido una igualdad exacta. En consonancia con esto, la afirmación del teorema 1 en este caso tendrá la forma

$$\mathbf{P}_\theta(\theta^- < \theta < \theta^+) \geqslant 1 - \varepsilon,$$

donde $\theta^\pm$ son las soluciones de las ecuaciones $G_\theta^{-1}(\varepsilon_1) = S$, $G_\theta^{-1}(1 - \varepsilon_2) = S$. Además, llamaremos intervalo confidencial de nivel $1 - \varepsilon$ el intervalo (θ^-, θ^+).

Si construimos el intervalo confidencial (θ^-, θ^+) con ayuda de la estimación θ^*, de la fig. 3 se deduce que este intervalo será tanto más estrecho cuanto más estrecho sea el intervalo $(G_\theta^{-1}(\varepsilon_1), G_\theta^{-1}(1 - \varepsilon_2))$ o bien, que es lo mismo, cuanto más concentrada sea la distribución de θ^* cerca de θ. Ahora bien, aquí llegamos al mismo problema que en la teoría de las estimaciones puntuales, o sea, a la determinación de las estimaciones θ^* que aprecian θ de la forma más exacta.

El problema relacionado con la construcción de los mejores intervalos confidenciales se examina más detalladamente en el § 3.8.

En vista de que la forma de las funciones de distribución $G_\theta(x)$ suele ser bastante compleja incluso para las familias simples de distribuciones citadas en el § 2, el referido procedimiento de inversión de $G_\theta(x)$ en la práctica resulta muy difícil. Por eso el cálculo de las fronteras confidenciales está considerablemente tubulado. En el ejemplo siguiente, donde ilustramos la construcción de los intervalos confidenciales según el esquema descrito en el teorema 1, para simplificar la exposición utilizaremos la aproximación normal.

Ejemplo 2. Sea $X \in \mathbf{B}_p$. En calidad de estimación para p tomemos la estimación eficiente $p^* = \nu/n$, donde ν es el número de casos favorables en n pruebas (el número ν puede designar, por ejemplo, la cantidad de artículos desechados que han sido descubiertos durante la verificación de control de n muestras. Es necesario construir el intervalo confidencial para la porción de artículos defectuosos p).

Tenemos ($q = 1 - p$)

$$G_p(x) = \mathbf{P}_p(p^* < x) = \mathbf{P}_p\left(\frac{\nu - np}{\sqrt{npq}} < \frac{xn - np}{\sqrt{npq}}\right).$$

Conforme al teorema 1 debemos resolver la ecuación

$$G_p(p^*) = \gamma \tag{9}$$

para los valores γ iguales a $\varepsilon/2$ y $1 - \varepsilon/2$. Cuando n son grandes, en virtud del teorema central del límite, $G_p(x) \approx \Phi((x - p)n/\sqrt{npq})$, donde $\Phi(y) = \Phi_{0,1}((-\infty, y))$, y, por consiguiente, la ecuación (9) puede ser sustituida por su aproximación

$$\Phi((p^* - p)n/\sqrt{npq}) \approx \gamma, \quad \gamma = \varepsilon/2,\ 1 - \varepsilon/2,$$

o bien, que es lo mismo, $|(p^* - p)n/\sqrt{nqq}| = \lambda_{\varepsilon/2} = \beta$,

$$(p^* - p)^2 = \beta^2 p(1 - p)/n.$$

Esta es la ecuación para las fronteras $p^{\pm}$ del intervalo confidencial, que no es otra cosa sino la ecuación de la elipse extendida para grandes n a lo largo de la bisectriz $p^* - p = 0$. Despejando p en esta ecuación, obtenemos

$$p^{\pm} \approx p^* \pm \beta\sqrt{p^*(1 - p^*)/n}.$$

No es difícil comprobar que obtendríamos ese mismo resultado si utilizáramos el enfoque asintótico expuesto en el punto 3.

Si n no es grande, conviene calcular $G_p(x)$ por la fórmula exacta

$$G_p(x) = \sum_{k < nx} C_n^k p^k (1 - p)^{n-k},$$

aplicando luego el procedimiento del teorema 1.

Supongamos, por ejemplo, que de $n = 10$ artículos $\nu = 2$ resultaron defectuosos. Entonces, cuando $\varepsilon = 0{,}05$, las fronteras exactas del intervalo confidencial son iguales a $p^- = 0{,}037$, $p^+ = 0{,}507$. La gran anchura del intervalo se explica por la poca información de que disponemos.

No obstante, si $n = 100$, $\nu = 20$, entonces, para $\varepsilon = 0{,}05$,

$$p^- = 0{,}137, \quad p^+ = 0{,}277.$$

Hemos tomado estas cifras de tablas especiales que dan la solución numérica del problema sobre los intervalos confidenciales para el número p, siendo diferentes n y ν (véase [8]).

5. Otros métodos de construcción de intervalos confidenciales. En este apartado examinaremos ciertas generalizaciones del procedimiento antes propuesto, relacionado con la construcción de intervalos confidenciales.

Teorema 2. *Admitamos que en* $\Theta \times \mathscr{X}^n$ *existe una función* $G(\theta, x)$, *tal, que la distribución* $\mathbf{H}(B) = \mathbf{P}_\theta(G(\theta, X) \in B)$ *no depende de* θ. *Supongamos también, que* $G(\theta, x)$, *para cada* x, *es continua y monotóna respecto a* θ.

Admitamos luego, que y^-, y^+ *satisfacen la relación* $\mathbf{H}((y^-, y^+)) = 1 - \varepsilon$. *Entonces las estadísticas*

$$\theta^- = G^{-1}(y^-, X),\ \theta^+ = G^{-1}(y^+, X),\ si\ G(\theta, \cdot)\uparrow,$$

y

$$\theta^- = G^{-1}(y^+, X),\ \theta^+ = G^{-1}(y^-, X),\ si\ G(\theta, \cdot)\downarrow,$$

son las fronteras del intervalo confidencial de nivel $1 - \varepsilon$. *Aquí* $G^{-1}(y, X)$ *es la solución de la ecuación* $G(\theta, X) = y$.

Demostración. En virtud de la monotonía de $G(\theta, x)$ (supongamos, para precisar, que $G(\theta, x)$ crece respecto a θ), el suceso $\{G^{-1}(y^-, X) < \theta < G^{-1}(y^+, X)\}$ coincide con el suceso $A = \{y^- < G(\theta, X) < y^+\}$.

Por definición de $\mathbf{H}(\cdot)$ e $y^\pm$ tenemos

$$\mathbf{P}_\theta(\theta^- < \theta < \theta^+) = \mathbf{P}_\theta(G^{-1}(y^-, X) < \theta < G^{-1}(y^+, X)) = \\ = \mathbf{P}_\theta(A) = \mathbf{H}((y^-, y^+)) = 1 - \varepsilon. \ \triangleleft$$

Observación 1. En el teorema 1, en calidad de $G(\theta, X)$ hemos examinado la función $G_\theta(S)$. Además se ha cumplido $\mathbf{H} = \mathbf{U}_{0,1}$.

Observación 2. Se puede examinar el análogo asintótico del teorema 2, admitiendo la existencia de la sucesión de funciones $\{G_n(\theta, x)\}$ continuas y monótonas respecto a θ y tales que, cuando $n \to \infty$,

$$\mathbf{P}_\theta(G_n(\theta, X) \in B) \to \mathbf{H}(B),$$

donde $\mathbf{H}(\cdot)$ no depende de θ. Entonces obtendremos el método de construcción de intervalos asintóticos confidenciales, que generaliza el método de construcción de intervalos asintóticos confidenciales mediante estimaciones asintóticamente normales, expuesto en el punto 3.

Ahora proponemos un método más (a la par con el teorema 1) de elección de la función $G(\theta, x)$ que figura en el teorema 2.

Teorema 3. *Sea* $F_\theta(x) = \mathbf{P}_\theta(x_1 < x)$, *con la particularidad de que*

1) $F_\theta(x)$ *es continua respecto a* x *para todos* $\theta \in \Theta$,

2) $F_\theta(x)$ *es continua y monótona respecto a* θ *para cualquier* x *registrada. Entonces la función*

$$G(\theta, x) = -\sum_{i=1}^{n} \ln(F_\theta(x_i))$$

satisface las condiciones del teorema 2.

Si los números $y^{\pm}$ *son tales que*

$$\frac{1}{\Gamma(n)} \int_{y^-}^{y^+} x^{n-1} e^{-x} dx = 1 - \varepsilon, \tag{10}$$

entonces $\theta^{\pm} = G^{-1}(y^{+}, X)$ *formarán las fronteras del intervalo confidencial de nivel* $1 - \varepsilon$.

Demostración. Verifiquemos el cumplimiento de las condiciones del teorema 2. Como, según la condición 1), $F_\theta(x_i)$ distribuida uniformemente en $[0, 1]$, entonces $-\ln F_\theta(x_i) \in \Gamma_{1,1}$ y $G(\theta, X) \in \Gamma_{1,n}$. Con otras palabras, $\mathbf{P}_\theta(G(\theta, X) \in B) = \Gamma_{1,n}(B)$ y $\mathbf{H} = \Gamma_{1,n}$ no depende de θ. La monotonía y la continuidad de $G(\theta, x)$ se deducen, para cada x, de la condición 2). Además, en virtud de (10)

$$\mathbf{H}((y^-, y^+)) = \Gamma_{1,n}((y^-, y^+)) = 1 - \varepsilon. \quad \triangleleft$$

También se pueden señalar algunas otras construcciones de los intervalos confidenciales. En este caso, al igual que en la teoría de estimación puntual, en seguida surge la pregunta acerca de qué intervalo confidencial debe considerarse el mejor si se han obtenido varios intervalos. En el § 3.8 trataremos de los enfoques que existen en este caso. Sin embargo, de la exposición precedente está claro que, de hecho, el problema de búsqueda del intervalo confidencial óptimo es en mucho muy parecido al problema de estimación puntual óptima. También está claro que si construimos los intervalos confidenciales utilizando las estimaciones puntuales, conviene dar preferencia a los intervalos confidenciales construidos con ayuda de las mejores estimaciones.

La semejanza de los problemas de optimación de las estimaciones puntual y por intervalo puede ser ilustrada citando el ejemplo de la afirmación siguiente.

Teorema 4. *Examinemos el intervalo asintótico confidencial* (θ^-, θ^+) *de nivel* $1 - \varepsilon$ *y supongamos que la variable aleatoria* $\theta^* = (\theta^+ + \theta^-)/2$ *es la estimación asintóticamente normal y asintóticamente central* (véase el punto 2 del § 25), *y la magnitud* $\Delta = (\theta^+ - \theta^-)/2$ *es tal, que* $\delta = \varliminf_{n\to\infty} \sqrt{n}\Delta$ *no depende de* X. *En este caso* $\delta \geqslant \beta/\sqrt{I(\theta)}$.

Esto quiere decir que la anchura del intervalo confidencial (θ^-, θ^+) no puede ser mucho menor que $2\beta/\sqrt{nI(\theta)}$, o sea, menor que la anchura del intervalo de nivel $1 - \varepsilon$ construido con ayuda de la e.v.m. $\hat{\theta}^*$.

Demostración. Admitamos lo contrario. Entonces habrá una subsucesión de los números $\{n'\}$ para los cuales $\Delta\sqrt{n'} \to c\beta/\sqrt{I(\theta)}$, $c < 1$. Como $\theta^{\pm} = \theta^* \pm \Delta$, entonces

$$1 - \varepsilon = \lim_{n'\to\infty} \mathbf{P}_\theta(\theta^- < \theta < \theta^+) = \lim_{n'\to\infty} \mathbf{P}_\theta(|\theta^* - \theta| < \Delta) =$$

$$= \lim_{n'\to\infty} \mathbf{P}_\theta(|\theta^* - \theta|\sqrt{n'} < c\beta/\sqrt{I(\theta)}) \leqslant \lim_{n\to\infty} \mathbf{P}_\theta(|\hat{\theta}^* - \theta|\sqrt{n} <$$

$$< c\beta/\sqrt{I(\theta)}). \qquad (11)$$

La última desigualdad se deduce del hecho de que la e.v.m. $\hat{\theta}^*$ es asintóticamente eficiente en la clase $\mathcal{K}^0$ de estimaciones asintóticamente centrales (véase el teorema 25.4). En vista de que en (11) el segundo miembro es menor que $1 - \varepsilon$, hemos obtenido la contradicción que demuestra el teorema. ◁

6. Caso multidimensional. El concepto de intervalo confidencial se generaliza en el caso del parámetro multidimensional $\theta \in R^k$ en el concepto de región confidencial o de conjunto confidencial.

Definición 4. El subconjunto aleatorio*) $\Theta^* = \Theta^*(\varepsilon, X)$ del espacio paramétrico Θ se llama *conjunto confidencial de nivel* $1 - \varepsilon$ si

$$\mathbf{P}_\theta(\Theta^* \ni \theta) \geqslant 1 - \varepsilon. \qquad (12)$$

Con otras palabras, el conjunto confidencial de nivel $1 - \varepsilon$ recubre el valor real desconocido de θ con una probabilidad no menor de $1 - \varepsilon$.

Definición 5. Si $X = [X_\infty]_n \in \mathbf{P}_\theta$, y si el conjunto aleatorio Θ^* satisface la relación

$$\lim_{n\to\infty} \inf \mathbf{P}_\theta(\Theta^* \ni \theta) \geqslant 1 - \varepsilon,$$

entonces Θ^* se llama *conjunto asintótico confidencial de nivel* $1 - \varepsilon$.

Los conjuntos confidenciales "exactos", incluso óptimos, se estudian en el § 8 del capítulo siguiente.

En lo que se refiere a los conjuntos asintóticos confidenciales, el principio de su construcción es el mismo de antes. Teniendo en cuenta el teorema 4, examinaremos a la vez los conjuntos confidenciales construidos con ayuda de la e.v.m. θ^*. Como sabemos, al cumplirse las condiciones (RR), $X \in \mathbf{P}_\theta$,

$$(\hat{\theta}^* - \theta)\sqrt{n}I^{1/2}(\theta) \in \Phi_{0,E}.$$

*) En este contexto diremos que el conjunto $\Theta^*(\varepsilon, X)$ es aleatorio si para cada t el conjunto $\{X: t \in \Theta^*(\varepsilon, X)\}$ es medible y, por lo tanto, también diremos que la probabilidad (12) está definida (compárese con el § 3.8.).

De aquí se deduce que

$$n(\hat{\theta}^* - \theta)I(\theta)(\hat{\theta}^* - \theta)^T \in \mathbf{H}_k,$$
$$n(\hat{\theta}^* - \theta)I(\hat{\theta}^*)(\hat{\theta}^* - \theta)^T \in \mathbf{H}_k.$$

Con otras palabras, si h_ε significa la cuantila de orden $1 - \varepsilon$ de la distribución χ^2 con k grados de libertad, entonces

$$\lim_{n\to\infty} \mathbf{P}_\theta(n(\theta - \hat{\theta}^*)I(\hat{\theta}^*)(\theta - \hat{\theta}^*)^T < h_\varepsilon) = 1 - \varepsilon. \tag{13}$$

Hemos construido el conjunto asintótico confidencial Θ^* de nivel $1 - \varepsilon$ que es un elipsoide cuyo centro se encuentra en el punto $\hat{\theta}^*$ y cuyos ejes se definen por la matriz $nI(\hat{\theta}^*)/h_\varepsilon$. En este caso no es obligatorio calcular la matriz $I(\theta)$ para la construcción de Θ^*. Como sabemos, al cumplirse las condiciones (RR), $X \in \mathbf{P}_\theta$,

$$L(X, \theta) - L(X, \hat{\theta}^*) \approx -\frac{n}{2}(\theta - \hat{\theta}^*)I(\hat{\theta}^*)(\theta - \hat{\theta}^*)^T.$$

Por eso el elipsoide Θ^* definido en (13) puede representarse como la población de los valores de θ para los cuales

$$L(X, \theta) - L(X, \hat{\theta}^*) \geqslant -h_\varepsilon/2.$$

En el § 28 hemos determinado que el límite de la $\mathbf{P}_\theta$-probabilidad de esta desigualdad (véase la observación 28.2) es igual a $1 - \varepsilon$.

De aquí resulta, en particular, que en el caso unidimensional, las fronteras $\theta^\pm$ del intervalo asintótico confidencial de nivel $1 - \varepsilon$ pueden ser definidas como las soluciones de la ecuación

$$L(X, \theta) - L(X, \hat{\theta}^*) = -h_\varepsilon/2 = -\beta^2/2.$$

§ 32. Distribuciones muestrales exactas e intervalos confidenciales exactos para poblaciones normales

Entre todas las distribuciones citadas en el § 2, la distribución normal tiene la mayor aplicación. Por eso en este párrafo examinaremos especialmente la construcción de los intervalos confidenciales para los parámetros α y σ^2 de la distribución Φ_{α,σ^2}.

1. Distribuciones exactas de las estadísticas $\bar{x}$, S_0^2. Supongamos que $X \in \Phi_{0,1}$ y que $C = \|c_{ij}\|$ $(i, j = 1, 2, \ldots, n)$ es una matriz ortogonal.

Examinemos la distribución del vector n-dimensional $Y = XC$, $Y = (y_1, \ldots, y_n)$, $y_i = \sum_{j=1}^{n} x_j c_{ji}$.

Lema 1. *Si C es una matriz ortogonal, entonces $Y \in \Phi_{0,1}$, o sea, las coordenadas $y_1, \ldots, y_n$ son variables aleatorias independientes, $y_i \in \Phi_{0,1}$, $i = 1, 2, \ldots, n$.*

Demostración. Sea t un vector $(t_1, \ldots, t_n)$. La normalidad de la distribución de X significa que su función característica es igual a

$$\mathbf{M}e^{itX^T} = e^{-\frac{1}{2}tmt^T},$$

donde $m = \|m_{ij}\|$ es una matriz de segundos momentos, que en nuestro caso es igual a la matriz unidad E para la cual $tEt^T = \sum_{j=1}^{n} t_j^2$,

$$\mathbf{M}e^{itX^T} = e^{-\frac{1}{2}\sum_{j=1}^{n} t_j^2}.$$

La función característica de la distribución compatible $y_1, \ldots, y_n$ (o de la distribución del vector Y) tiene la forma

$$f(t) = \mathbf{M}e^{itY^T} = \mathbf{M}e^{itC^TX^T}.$$

Sustituyendo las variables $t = uC$ y notando que $CC^T = E$, obtenemos

$$f(t) = \mathbf{M}e^{iuCY^T} = \mathbf{M}e^{iuX^T} = e^{-\frac{1}{2}\sum_{j=1}^{n} u_j^2} = e^{-\frac{1}{2}\sum_{j=1}^{n} t_j^2}.$$

Esto quiere decir que Y tiene la misma función característica y, por lo tanto, la misma distribución que X. ◅

Ahora demostremos una afirmación llamada lema de Fisher, que es muy importante para la exposición ulterior.

Lema 2. *Supongamos, como antes, que* $X \in \Phi_{0,1}$, *que* C *es una matriz ortogonal y que* $Y = (y_1, \ldots, y_n) = XC$. *Entonces, la forma cuadrática*

$$T(X) = \sum_{i=1}^{n} x_i^2 - y_1^2 - \ldots - y_r^2$$

no depende de las variables aleatorias $y_1, \ldots, y_r$ *y tiene una distribución* χ^2 *con* $n - r$ *grados de libertad*:

La demostración es casi evidente, ya que después de aplicar la transformación ortogonal de C, obtenemos

$$T(X) = \sum_{i=1}^{n} y_i^2 - y_1^2 - \ldots - y_r^2 = y_{r+1}^2 + \ldots + y_n^2.$$

Solamente queda utilizar el lema 1. ◅

Pasemos ahora al estudio de la distribución compatible de las estadísticas $\bar{x}$ y $S_0^2 = \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2$.

Teorema 1. *Sea* $X \in \Phi_{\alpha,\sigma^2}$. *Entonces*

1) $(\bar{x} - \alpha)\sqrt{n}/\sigma \in \Phi_{0,1}$,

2) $(n-1)S_0^2/\sigma \in \mathbf{H}_{n-1}$,

3) *las variables aleatorias* $\bar{x}$ y S_0^2 *son independientes.*

Demostración. La afirmación 1 es evidente. Además está claro que sin limitar la generalidad podemos considerar $\alpha = 0$, $\sigma = 1$. Tenemos

$$(n-1)S_0^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2.$$

Notemos que

$$\sqrt{n}\bar{x} = \frac{1}{\sqrt{n}} x_1 + \ldots + \frac{1}{\sqrt{n}} x_n$$

y que el vector columna n-dimensional $\begin{matrix} 1/\sqrt{n} \\ \vdots \\ 1/\sqrt{n} \end{matrix}$ (su norma vale 1) siempre puede ser completado hasta cualquier matriz ortogonal C. Entonces $y_1 = \sqrt{n}\bar{x}$ es la primera coordenada $Y = XC$ y, en virtud del lema 2, obtenemos que

$$(n-1)S_0^2 = \sum_{i=1}^{n} x_i^2 - y_1^2 \in \mathbf{H}_{n-1}$$

y que las variables aleatorias $(n-1)S_0^2$ e $y_1 = \sqrt{n}\bar{x}$ son independientes. ◁

Corolario 1. *Sea $X \in \Phi_{\alpha,\sigma^2}$. Entonces la variable aleatoria $t = (\bar{x} - \alpha)\sqrt{n}/S_0 \in \mathbf{T}_{n-1}$, o sea, tiene una distribución de Student con $n-1$ grados de libertad.*

Esto se deduce del teorema 1 y de la representación

$$t = \frac{(\bar{x}-\alpha)\sqrt{n}}{\sigma} \cdot \frac{1}{\sqrt{\frac{1}{n-1} \cdot \frac{(n-1)S_0^2}{\sigma^2}}}. \quad ◁$$

La afirmación del teorema 1 acerca de la independencia de S_0^2 y $\bar{x}$ puede ser amplificada. Resulta que $\bar{x}$ no depende del vector $X - \bar{x}$ (o sea, que no depende de los sumandos de S_0^2). Esto se deduce de la normalidad de $\bar{x}$ y de $X - \bar{x}$, así como de la no correlatividad de las variables aleatorias $\bar{x}$ y $x_i - \bar{x}$, la cual se desprende de la igualdad ($\alpha = 0$)

$$\mathbf{M}(x_1 - \bar{x})\bar{x} = \frac{1}{n^2}\left[(n-1)\mathbf{M}x_1^2 - \mathbf{M}\left(\sum_{i=2}^{n} x_i\right)^2\right] = 0.$$

2. Construcción de intervalos confidenciales exactos para los parámetros de distribución normal. Examinemos primeramente dos situaciones elementales.

a) Supongamos que $X \in \Phi_{\alpha,\sigma}$, y que σ^2 se conoce. Es preciso construir el intervalo confidencial de nivel $1 - \varepsilon$ para el parámetro α. En este caso la forma del intervalo confidencial se deduce, evidentemente, de las igualdades

$$\mathbf{P}(|\bar{x} - \alpha)\sqrt{n}/\sigma| < \beta) = \mathbf{P}(-\sigma\beta/\sqrt{n} < \bar{x} - \alpha < \sigma\beta/\sqrt{n} = 1 - \varepsilon,$$

donde, como antes, $\beta = \lambda_{\varepsilon/2}$, $\Phi_{0,1}((-\infty, \lambda_\delta)) = 1 - \delta$, así que

$$\alpha^{\pm}(\varepsilon, X) = \bar{x} \pm \sigma\beta/\sqrt{n}.$$

Proponemos que el lector, en forma de ejercicio, haga uso de un procedimiento un poco más formal, expuesto en el teorema 31.2, con la utilización de la función $G(\alpha, X) = (\bar{x} - \alpha)\sqrt{n}/\sigma \in \Phi_{0,1}$.

b) Ahora supongamos que se conoce α. Es necesario construir el intervalo confidencial de nivel $1 - \varepsilon$ para σ^2.

Pongamos

$$S_1^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \alpha)^2.$$

Es evidente que en este caso $nS_1^2/\sigma^2 \in \mathbf{H}_n$ y, por consiguiente,

$$\mathbf{P}(y_n^- < nS_1^2/\sigma^2 < y_n^+) = \mathbf{H}_n((y_n^-, y_n^+)) = \mathbf{P}(nS_1^2/y_n^+ < \sigma^2 < nS_1^2/y_n^-).$$

Ahora bien, las fronteras del intervalo confidencial de nivel $1 - \varepsilon$ tendrán la forma

$$(\sigma^2)^{\pm} = nS_1^2/y_n^{\pm}$$

para todos $y_n^{\pm}$ tales que $\mathbf{H}_n((y_n^-, y_n^+)) = 1 - \varepsilon$.

Si se utiliza el procedimiento del teorema 31.2, conviene poner $G(\sigma, X) = nS_1^2/\sigma^2 \in \mathbf{H}_n$.

Pasemos ahora al caso cuando ambos parámetros α y σ^2 se desconocen.

c) Con el fin de construir el intervalo confidencial para σ^2, hagamos uso de la estadística $G_1(\sigma, X) = (n - 1)S_0^2/\sigma^2$. En virtud del teorema 1, $G_1(\sigma, X) \in \mathbf{H}_{n-1}$. Luego procedemos al igual que en el caso b). Las fronteras del intervalo confidencial para σ^2 tendrán la forma

$$(\sigma^2)^{\pm} = (n - 1)S_0^2/y_{n-1}^{\pm}.$$

Es fácil ver que las estadísticas $G(\sigma, X)$ y $G_1(\sigma, X)$ en los casos b) y c) tienen la misma distribución y, por lo tanto, dan los mismos intervalos confidenciales para σ^2 siempre que en el caso b) tengamos una observación más que en el caso c). Hablando figuradamente, en el caso c) "perdemos" una observación debido a la existencia de una indeterminación adicional,

o sea, del parámetro desconocido α. Esta observación se destina, en cierto sentido, a estimar el parámetro "obstaculizante"*) α.

d) Construyamos ahora el intervalo confidencial para α. Hagamos uso de la estadística $G_1(\alpha, X) = (\bar{x} - \alpha)\sqrt{n}/S_0$. En virtud del corolario del teorema 1,

$$G_1(\alpha, X) \in \mathbf{T}_{n-1}.$$

En vista de que la función $G_1(\alpha, X)$ satisface las condiciones del teorema 31.2, los razonamientos ulteriores repiten exactamente los correspondientes razonamientos en los casos a), b) y c). Las fronteras del intervalo confidencial tienen la forma (para simplificar la exposición tomamos un intervalo simétrico)

$$\alpha^{\pm} = \bar{x} \pm \tau_\varepsilon S_0/\sqrt{n},$$

donde τ_ε se determina de la igualdad

$$\mathbf{P}(|t_{n-1}| < \tau_\varepsilon) = \mathbf{T}_{n-1}((-\tau_\varepsilon, \tau_\varepsilon)) = 1 - \varepsilon.$$

Nótese que si el valor de S_0 es próximo al de σ, entonces el intervalo confidencial obtenido será más ancho que el dado en a), ya que $\tau_\varepsilon > \beta$ (véase la observación en el § 2). Esto se explica, como antes, por la existencia del parámetro "obstaculizante" σ, el cual se conoce en a).

Los números $y^{\pm}$, para los cuales en las investigaciones citadas se ha cumplido la relación

$$\mathbf{P}(G(\theta, X) \in (y^-, y^+)) = 1 - \varepsilon,$$

en la práctica suelen determinarse con ayuda de las tablas de la estadística matemática.

En el § 3.8 mostraremos que los intervalos confidenciales construidos en este párrafo son, desde cierto punto de vista, los mejores.

---

*) Es interesante notar que, a pesar de las ideas intuitivas iniciales, por una observación $x_1 \in \Phi_{\alpha,\sigma^2}$ es posible construir el intervalo confidencial para σ^2, siendo α desconocido. Los siguientes razonamientos que muestran esto fueron comunicados a nosotros por L. N. Bolshakov.

Escojamos u de modo que $\Phi(1/u) - \Phi(-1/u) = \varepsilon$, donde $\Phi(x) = \Phi_{0,1}((-\infty, x))$. Entonces

$$\mathbf{P}(\sigma > u|x_1|) = \mathbf{P}(-\sigma/u < x_1 < \sigma/u) = \mathbf{P}\left(-\frac{1}{u} - \frac{\alpha}{\sigma} < \frac{(x_1 - \alpha)}{\sigma} < \frac{1}{u} - \frac{\alpha}{\sigma}\right) =$$

$$= \Phi\left(\frac{1}{u} - \frac{\alpha}{\sigma}\right) - \Phi\left(-\frac{1}{u} - \frac{\alpha}{\sigma}\right) \leq \Phi\left(\frac{1}{u}\right) - \Phi\left(-\frac{1}{u}\right) = \varepsilon. \lhd$$

CAPÍTULO 3

Teoría de verificación de las hipótesis

En los §§ 1—3, 11 se expone la teoría de verificación de un número finito (en particular, dos) de hipótesis simples.

Los §§ 4—12 están dedicados a los métodos de construcción de criterios óptimos para verificar dos hipótesis compuestas. En particular, se examinan los criterios bayesianos y minimax (los §§ 4 y 9) y se utilizan los principios de suficiencia, de carácter no desplazable y de invariación para construir los criterios uniformemente más potentes.

En los §§ 13—17 se estudian los métodos de construcción de criterios asintóticamente óptimos.

§ 1. Verificación de un número finito de hipótesis simples

1. Planteamiento del problema. Concepto de criterio estadístico. Criterio más potente. En este capítulo se tratará de la verificación de cualesquiera suposiciones (hipótesis) respecto a la distribución **P** de la cual se ha extraído la muestra X. Aquí, al igual que en la teoría de las estimaciones, no existiría tal problema, si la distribución **P**, de la cual se extrae la muestra X, fuera conocida.

La decisión de que es cierta o no la hipótesis dada H debe basarse exclusivamente en el conocimiento de la muestra $X \in \mathbf{P}$ extraída y, posiblemente, también en el conocimiento de la información a priori respecto a **P** si disponemos de ella.

Ahora bien, para determinar el procedimiento de toma de decisión basándonos en la muestra X, debemos establecer, de una u otra forma, la aplicación del espacio muestral $\mathscr{X}^n$ en el conjunto de hipótesis que se examinan. Tal aplicación suele llamarse criterio estadístico. Las definiciones exactas para diferentes situaciones concretas se darán más adelante.

Comencemos por el problema más simple: verificación de un número finito de hipótesis simples.

Definición 1. Llamaremos *hipótesis simple* cualquier suposición que defina unívocamente la distribución de la muestra X.

Supongamos que se dan r distribuciones $\mathbf{P}_1, \ldots, \mathbf{P}_r$, y supongamos que sabemos que X es la muestra de una de estas distribuciones. El problema consiste en determinar a qué $\mathbf{P}_j$ precisamente, $j = 1, 2, \ldots, r$, pertenece X. Cada r hipótesis

$$H_j = \{X \in \mathbf{P}_j\} \tag{1}$$

será simple y, por consiguiente, se tratará de la verificación de r hipótesis simples.

En este capítulo, al igual que en el capítulo 2, examinaremos con frecuencia el caso paramétrico cuando la muestra X se ha extraído de la distribución $\mathbf{P}_\theta \in \mathscr{P} = \{\mathbf{P}_\theta\}_{\theta \in \Theta}$. En este caso, al cumplirse las condiciones (A_0), las hipótesis simples se escribirán en la forma: $H_j = \{X \in \mathbf{P}_{\theta_j}\}$, donde $\theta_1, \ldots, \theta_r$ son los puntos fijos de Θ. El caso (1) también puede considerarse como paramétrico con un conjunto finito $\Theta = \{1, \ldots, r\}$.

Estos razonamientos muestran que no hay una diferencia de principio entre el problema de estimación de los parámetros y el problema de verificación de las hipótesis: en ambos casos determinamos el valor desconocido de θ. Sin embargo, existe cierta diferencia y ésta consiste en que en el problema de verificación de las hipótesis, los valores posibles de θ son discretos, y los enfoques relacionados con la comparación, digamos, de las desviaciones estándar, desarrollados en el capítulo 2, aquí son inaplicables. En este caso escogeremos otros criterios para comparar las reglas de aceptación de unas u otras hipótesis, basándonos en la muestra X.

Con el carácter discreto del conjunto de los posibles valores de θ también está relacionada otra nueva cualidad que aparece aquí: ahora podemos, con una probabilidad no nula, indicar exactamente el valor desconocido de θ_l (o la distribución $\mathbf{P}_{\theta_l}$), mientras que en los problemas de estimación de los parámetros, la probabilidad de tal suceso es, por regla general, igual a cero.

Definición 2. Se llama *criterio estadístico* para verificar r hipótesis $H_1, \ldots, H_r$ toda aplicación medible $\delta: \mathscr{X}^n \to \{H_1, \ldots, H_r\}$.

En otros términos, $\delta(X)$ es una "variable" aleatoria que toma los valores $H_1, H_2, \ldots, H_r$: si $\delta(X) = H_k$, entonces aceptamos la hipótesis H_k (o sea, consideramos que $\theta = \theta_k$ en el caso paramétrico).

La aplicación $\delta(\cdot)$ se llama, a veces, *regla de decisión* o *función de decisión*. Claro está que la asignación de la regla de decisión es equivalente a la partición del espacio $\mathscr{X}^n$ en r conjuntos borelianos $\Omega_1, \Omega_2, \ldots, \Omega_r$ disjuntos, en los cuales se aceptan las hipótesis $H_1, H_2, \ldots, H_r$, respectivamente.

La calidad del criterio se caracteriza, con más frecuencia, por el conjunto de probabilidades de decisiones erróneas:

$$\alpha_i = \alpha_i(\delta) = \mathbf{P}_i(X \notin \Omega_i) = \mathbf{P}_i(\delta(X) \neq H_i).$$

El número α_i es la probabilidad de rechazar la hipótesis H_i cuando ésta es cierta. Este número se denomina *probabilidad del error de i-ésimo género del criterio δ.*

Si logramos escoger el criterio δ de modo que todos los números α_i sean pequeños, entonces, según nuestro principio fundamental mencionado en el § 2.31, consideraremos que en una sola prueba el error es prácticamente imposible y declararemos que es cierta la hipótesis H_k si $\delta(X) = H_k$. En este caso nos equivocaremos, aproximadamente, en parte de los casos $\alpha_i = \mathbf{P}_i(\delta(X) \neq H_i)$ si en realidad es cierta H_i.

Es deseable, desde luego, efectuar la verificación de las hipótesis de modo que se reduzca al mínino la probabilidad de todos los errores. No obstante, si se establece el volumen de la muestra X, entonces no podremos dirigir simultáneamente todas las probabilidades de los errores. Se puede sólo, fijando algunas de las probabilidades de errores, tratar de minimizar las demás.

Aquí llegamos a la cuestión de cómo comparar entre sí diferentes criterios. Introduzcamos en el conjunto de todos los criterios, para verificar las hipótesis $H_1, \ldots, H_r$, un orden parcial.

Definición 3. El criterio δ_1 es mejor que el δ_2 si para todos $i = 1, 2, \ldots, r$

$$\alpha_i(\delta_1) \leqslant \alpha_i(\delta_2)$$

y al menos para un i tiene lugar la desigualdad estricta.

Sin embargo, los criterios δ_1 y δ_2 no siempre, ni mucho menos, pueden compararse desde este punto de vista. Al igual que pueden ser incomparables dos estimaciones θ_1^* y θ_2^* desde el punto de vista del enfoque estándar, cuando en calidad de criterio tomamos $\mathbf{M}_\theta(\theta^* - \theta)^2$. Para tener la posibilidad de comparar los criterios es necesario contraer el conjunto de las reglas de decisión que se examinan. Para esto examinemos las clases

$$K_{\alpha_1,\ldots,\alpha_{r-1}} = \{\delta\colon \alpha_j(\delta) = \alpha_j;\quad j = 1, 2, \ldots, r-1\}.$$

En las clases $K_{\alpha_1,\ldots,\alpha_{r-1}}$ ya se puede establecer la relación de orden entre los criterios en cuanto a la magnitud α_r: cuanto menor sea $\alpha_r(\delta)$, tanto mejor será el criterio.

Definición 4. El criterio $\delta_0 \in K_{\alpha_1,\ldots,\alpha_{r-1}}$ se llama *criterio más potente* (*c.m.p.*) *en la clase* $K_{\alpha_1,\ldots,\alpha_{r-1}}$ si para cualquier $\delta \in K_{\alpha_1,\ldots,\alpha_{r-1}}$,

$$\alpha_r(\delta_0) \leqslant \alpha_r(\delta).$$

Recordemos que hemos hecho algo semejante en el capítulo 2 al compa-

rar las estimaciones. Allí hemos destacado, por ejemplo, las clases K_b de estimaciones con desplazamiento registrado.

A la par con el enfoque recién introducido en la teoría de verificación de las hipótesis, al igual que en la teoría de estimaciones, existen otros dos enfoques que permiten ordenar el conjunto de todas las reglas de decisión con ayuda de una sola característica numérica: son los enfoques bayesiano y minimax.

Antes de estudiar los métodos de construcción de los criterios más potentes en las clases $K_{\alpha_1,\ldots,\alpha_{r-1}}$, examinemos estos dos enfoques.

2. Enfoque bayesiano. Este enfoque supone que la distribución $\mathbf{P}_j$ de la que fue extraída la muestra X se ha elegido aleatoriamente. En este caso las hipótesis $H_j = \{X \in \mathbf{P}_j\}$, $j = 1, \ldots, r$ serán sucesos aleatorios, y designaremos las probabilidades de estos sucesos por

$$\mathbf{Q}(H_j) = q(j),$$

así que $\mathbf{Q}$ es una distribución a priori en el conjunto de las hipótesis $\{H_1, \ldots, H_r\}$, y $q(j)$ son las probabilidades a priori de dichas hipótesis (compárese con el § 2.11). En este caso es más fácil comparar los criterios, puesto que aquí podemos determinar la probabilidad media $\alpha_Q(\delta)$ del error del criterio δ:

$$\alpha_Q(\delta) = \sum_{j=1}^{r} \mathbf{Q}(H_j)\mathbf{P}_j(\delta(X) \neq H_j) = \sum_{j=1}^{r} q(j)\alpha_j(\delta), \tag{2}$$

y de este modo ordenar por completo el conjunto de criterios en cuanto a la magnitud $\alpha_Q(\delta)$.

Definición 5. El criterio $\delta = \delta_Q$ que minimiza la probabilidad del error $\alpha_Q(\delta)$ se denomina *criterio bayesiano correspondiente a la distribución a priori* $\mathbf{Q}$.

Supongamos que se cumple la condición (A_μ), o sea, las distribuciones $\mathbf{P}_j$ tienen densidades $f_j(x)$ respecto a cierta σ-finita medida μ. Al igual que antes, la función $f_j(X) = \prod_{j=1}^{n} f_j(x_j)$ se llamará *función de verosimilitud.*

La función $f(x) = \Sigma q(j)f_j(x)$ es la densidad incondicional de la distribución de X respecto a la medida μ^n, y $q(j)f_j(x)$ es la densidad de la distribución compatible del par (θ, X) en el que el número θ de la hipótesis se elige al azar.

Ahora bien, si se da la muestra X, entonces, en el caso bayesiano se puede construir la distribución a posteriori $\mathbf{Q}_x$ de las hipótesis H_j (la medida λ que figura en el § 2.11, aquí es una medida de cálculo) la cual se determina por la fórmula de Bayes:

$$\mathbf{Q}_X(H_k) = q(k/X) = \frac{q(k)f_k(X)}{f(X)}. \tag{3}$$

Esta es la *distribución condicional de θ respecto a X.*

Por **M** designaremos la esperanza matemática incondicional que corresponde a la distribución **P** del par (θ, X).

Teorema 1. 1) *La probabilidad del error* $\alpha_Q(\delta)$ *de cualquier criterio* δ *satisface la desigualdad*

$$\alpha_Q(\delta) \geqslant 1 - \mathbf{M} \max_j q(j/X). \tag{4}$$

2) *Para que el criterio* $\delta = \delta_Q$ *sea bayesiano para la distribución a priori* **Q**, *es necesario y suficiente que para* **P** *de casi todos los valores de* X, *este criterio satisfaga las relaciones*

$$\delta(X) = H_k \quad \text{si} \quad q(k/X) = \max_j q(j/X). \tag{5}$$

Para $\delta = \delta_Q$ *en la desigualdad* (4) *se alcanza la igualdad.*

Nótese que el segundo miembro en (4) no depende de δ.

Demostración. Supongamos que se da el criterio δ. Examinemos el suceso D_δ que consiste en que el criterio δ conduce a la decisión errónea:

$$D_\delta = \bigcup_{j=1}^{r} \{\theta = j,\ \delta(X) \neq H_j\}.$$

Entonces, evidentemente que $\alpha_Q(\delta) = \mathbf{P}(D_\delta)$ y la notación (2) será el resultado obtenido al promediar sucesivamente: primero respecto a X al ser registrado $\theta = j$, y luego respecto a θ. Pero también podemos escribir $\alpha_Q(\delta)$ de otro modo: primero promediar respecto a θ, siendo registrado (X), y luego respecto a X:

$$\alpha_Q(\delta) = \int \mathbf{P}(D_\delta/X = x) f(x) \mu(dx) =$$
$$= \mathbf{MP}(D_\delta/X) = \mathbf{M} \sum_{j=1}^{r} \mathbf{P}(\theta = j,\ \delta(X) \neq H_j/X).$$

Pero $\delta(X)$ es medible respecto a X, por eso

$$\mathbf{P}(\theta = j,\ \delta(X) \neq H_j/X) = I_{\{\delta(X) \neq H_j\}} \mathbf{P}(\theta = j/X) = (1 - I_{\{\delta(X) = H_j\}}) q(j/X).$$

De aquí obtenemos

$$\alpha_Q(\delta) = 1 - \mathbf{M} \sum_{j=1}^{r} q(j/X) I_{\{\delta(X) = H_j\}} \geqslant 1 - \mathbf{M} \max_j q(j/X).$$

La primera afirmación del teorema queda demostrada.

La suficiencia de la segunda afirmación del teorema se deduce con evidencia de la primera, ya que la frontera inferior establecida para $\alpha_Q(\delta)$ se alcanza para el criterio δ_Q definido en (5). La modificación de $\delta_Q(X)$ en el conjunto de **P**-probabilidad nula, por lo visto no modifica $\alpha_Q(\delta_Q)$.

La necesidad de la segunda afirmación se demuestra de manera igualmente sencilla. En efecto, supongamos que $\delta = \delta_Q$ es el criterio bayesiano y que $\delta(X) = H_k$, $q(k/X) < q(l/X) = \max_j q(j/X)$ para $X \in A$, $\mathbf{P}(A) > 0$. Entonces, para el criterio $\delta_1(X)$, que se distingue de $\delta(X)$ sólo en el conjunto A: $\delta_1(X) = H_l$ para $X \in A$, obtenemos

$$\mathbf{P}(D_{\delta_1}; A) = \mathbf{P}(A) - \mathbf{M}\left[\sum_j q(j/X) I_{\{\delta_1(X) = H_j\}}; A\right] =$$

$$= \mathbf{P}(A) - \mathbf{M}[q(l/X); A] < \mathbf{P}(A) - \mathbf{M}[q(k/X); A] = \mathbf{P}(D_\delta; A);$$

$$\mathbf{P}(D_{\delta_1}) < \mathbf{P}(D_\delta) = \mathbf{P}(D_{\delta_Q}).$$

Hemos obtenido la contradicción. ▵

Cabe señalar que la notación (5) aún no define por completo el criterio δ_Q: ella no aclara bien qué hipótesis deben aceptarse cuando resultaron máximos dos o más valores de $q(j/X)$. Se trata, evidentemente, de la definición de la función $\delta_Q(X)$ en las fronteras

$$\Gamma_k = \{x \in \mathscr{X}^n:\ q(k)f_k(x) = \max_{j \neq k} q(j)f_j(x)\}$$

de los conjuntos

$$\tilde{\Omega}_k^Q = \{x \in \mathscr{X}^n:\ q(k)f_k(x) > \max_{j \neq k} q(j)f_j(x)\} \tag{6}$$

en los cuales, según (5), como criterio δ_Q se toma la hipótesis H_k.

Por consiguiente, $\tilde{\Omega}_k^Q$ es el "interior" de la región

$$\Omega_k^Q = \{x \in \mathscr{X}^n:\ \delta_Q(x) = H_k\}$$

de aceptación de la hipótesis H_k y necesitamos, en adición a (6), determinar tan sólo qué puntos de la frontera Γ_k pertenecen y no pertenecen a Ω_k^Q. Pero este problema, como se deduce de los razonamientos citados, puede ser resuelto muy sencillamente: podemos asociar los puntos de Γ_k a cualquiera de las regiones "adyacentes" $\tilde{\Omega}_j^Q$; en este caso obtenemos el mismo valor de $\alpha_Q(\delta)$, puesto que (5) será cumplida. Mejor dicho, si $A \subset \subset \Gamma_{k_1} \cap \ldots \cap \Gamma_{k_l}$, entonces para $X \in A$, según el criterio bayesiano, no importa cuál de las hipótesis $H_{k_1}, \ldots, H_{k_l}$ será aceptada. Podemos incluso tomar la decisión al azar, o sea, con probabilidad p_{k_i} elegir la hipótesis H_{k_i}, $i = 1, \ldots, l$, $\sum_{i=1}^{l} p_{k_i} = 1$. En este caso el valor de $\alpha_Q(\delta)$ no variará.

Aquí llegamos a un concepto más general del criterio estadístico randomizado (de la palabra inglesa *random*) que resulta muy útil.

Definición 6. Se llama *criterio estadístico randomizado*, para comprobar las hipótesis $H_1, \ldots, H_r$, cualquier aplicación medible $\pi: \mathscr{X}^n \to R^{(r)}$, donde $R^{(r)}$ es el conjunto de vectores $(\pi_1, \ldots, \pi_r)$, $\pi_l \geq 0$, $\sum_{l=1}^{r} \pi_l = 1$.

El criterio randomizado, a cada $x \in \mathscr{X}^n$ le pone en correspondencia con la distribución de las probabilidades $\pi(x) = (\pi_1(x), \ldots, \pi_r(x))$ en el conjunto $\{H_1, \ldots, H_r\}$, y la decisión final acerca de la aceptación de la hipótesis "se sortea" al azar con esta distribución ya independientemente de X, después de haber determinado $\pi_i(X)$.

El criterio estadístico ordinario es, evidentemente, un caso particular del randomizado, cuando todos π_i equivalen a 0 y sólo uno es igual a 1. Tales criterios adquirieron el nombre de criterios *no randomizados.*

El error de i-ésimo género $\alpha_i(\pi)$ para el criterio randomizado se determina análogamente:

$$\alpha_i(\pi) = \mathbf{P}_i \text{ (no aceptar } H_i) = 1 - \mathbf{M}_i\pi_i(X).$$

En el caso bayesiano, el problema de minimización

$$\alpha_Q(\pi) = \sum_{j=1}^{r} q(j)\alpha_j(\pi)$$

se examina de manera absolutamente semejante. Si, como antes, designamos por θ el número de la hipótesis elegida al azar, con una distribución a priori $\mathbf{Q}$, de modo que $\mathbf{Q}(\theta = j) = q(j)$, y por $\mathbf{M}$, también como antes, designamos el símbolo de la esperanza matemática incondicional, entonces

$$\alpha_Q(\pi) = 1 - \sum_{j=1}^{r} q(j)\mathbf{M}_i\pi_i(X) = 1 - \mathbf{M}\pi_\theta(X) = 1 - \mathbf{MM}(\pi_\theta(X)/X) =$$

$$= 1 - \mathbf{M}\sum_{j=1}^{r} q(j/X)\pi_j(X) \geqslant 1 - \mathbf{M}\max_j q(j/X).$$

Así pues, hemos obtenido la misma frontera inferior tanto para $\alpha_Q(\pi)$ como para los criterios no randomizados. Esto significa que ampliando la clase de criterios, en nuestro caso no podemos mejorar el valor de $\alpha_Q(\delta)$. Es más, el valor mínimo se alcanza en el criterio no randomizado δ_Q. Sin embargo, en este caso el número de criterios *randomizados* bayesianos π^Q, o sea, de criterios para los cuales $\alpha_Q(\pi^Q) = \alpha_Q(\delta_Q)$, será mucho mayor que los no randomizados, ya que en el conjunto

$$\Gamma_{k_1,\ldots,k_l} = \bigcap_{i=1}^{l} \Gamma_{k_i} \bigcap_{j \neq k_1,\ldots,k_l} \bar{\Gamma}_j,$$

donde $\bar{\Gamma} = \mathscr{X}^n \setminus \Gamma$, podemos tomar, en calidad de $\pi^Q(x)$, cualquier vector del subconjunto $R_{k_1,\ldots,k_l} \subset R^{(r)}$ compuesto de vectores π en los cuales sólo se diferencian del cero las coordenadas con números $k_1, \ldots, k_l$. Es evidente que R_k se compone del único vector e_k en el que la k-ésima coordenada es igual a 1, y las demás, a cero, y debemos poner

$$\pi^Q(x) = e_k \quad \text{cuando} \quad x \in \tilde{\Omega}_k^Q.$$

Como las relaciones expuestas con una exactitud de hasta los valores de $\pi^Q(x)$ en el conjunto de P-medida 0, son necesarias y suficientes para que

$$\alpha_Q(\pi^Q) = \alpha_Q(\delta_Q) = 1 - \mathbf{M} \max_j q(j/X),$$

podemos, a la par con el teorema 1, enunciar la afirmación siguiente:

Teorema 1A. 1) *Para cualquier criterio randomizado,*

$$\alpha_Q(\pi) \geqslant 1 - \mathbf{M} \max_j q(j/X).$$

2) *Para que el criterio π^Q sea bayesiano es necesario y suficiente el cumplimiento de las relaciones*

$$\pi^Q(x) = e_k \quad \textit{cuando} \quad x \in \tilde{\Omega}_k^Q, \tag{7}$$
$$\pi^Q(x) \in R_{k_1,\ldots,k_l} \quad \textit{cuando} \quad x \in \Gamma_{k_1,\ldots,k_l}$$

para $\mathbf{P}$ *de casi todos los valores de* x.

3) *Para todos* $g_j \geqslant 0$, $j = 1, \ldots, r$; $\sum_{j=1}^{r} g_j = 1$ *es válida la desigualdad*

$$\alpha_Q(\pi^Q) = \sum_{j=1}^{r} q(j)\alpha_j(\pi^Q) \leqslant \sum_{j=1}^{r} q(j)(1 - g_j). \tag{8}$$

Si $\min_j q_j > 0$ *y no todos* $f_i(x)$ *coinciden, o sea, si existen los valores* k, j *y el conjunto* A, $\mathbf{P}(A) > 0$ *en el que* $f_k(x) \neq f_j(x)$, *entonces el signo en la desigualdad* (8) *será estricto.*

Observación 1. De (8) se deduce que

$$\alpha_Q(\pi^Q) \leqslant 1 - \max_j q(j). \tag{9}$$

Aquí en el segundo miembro figura la probabilidad del error del criterio que elige H_k si $q(k) = \max_j q(j)$ (éste es el criterio bayesiano entre todos los criterios no dependientes de la muestra X).

Demostración del teorema 1A. Ya hemos demostrado las dos primeras afirmaciones. Para demostrar la última afirmación es suficiente comparar el criterio bayesiano π^Q con el criterio $\pi^\circ(X) = g = (g_1, \ldots, g_r)$ no dependiente de X y para el cual, como es evidente, $\alpha_j(\pi^\circ) = 1 - g_j$,

$$\alpha_Q(\pi^\circ) = \sum_{j=1}^{r} q(j)(1 - g_j) \geqslant \alpha_Q(\pi^Q).$$

Si en (8) tiene lugar la desigualdad, entonces el criterio $\pi^\circ(X) = g = \text{const}$ será bayesiano. Según la segunda afirmación del teorema, esto es posible únicamente en el caso cuando $q(1/X) = \ldots = q(r/X)$ $\mathbf{P}$ casi por doquier. Esto, a su vez, es posible únicamente cuando $f_1(X) = \ldots = f_r(X)$ $\mathbf{P}$ casi por doquier, $q(1) = \ldots = q(r)$. ◁

Así pues, la introducción de los criterios randomizados no permite disminuir la probabilidad del error de α_Q, pero aumenta la propia variedad de los criterios y, en particular, el número de criterios bayesianos π^Q. Esta circunstancia resulta, a veces, útil.

En lo sucesivo, por criterio estadístico entenderemos, por regla general, el criterio randomizado π.

3. Enfoque minimax. Mientras en el caso bayesiano hemos medido la calidad del criterio según la magnitud media $\alpha_Q(\pi) = \Sigma q(j)\alpha_j(\pi)$, ahora compararemos los valores máximos

$$\alpha(\pi) = \max_j \alpha_j(\pi) = \max_Q \alpha_Q(\pi).$$

Es evidente que esto también permite ordenar el conjunto de todos los criterios.

Definición 7. El criterio $\pi = \bar{\pi}$ para el cual

$$\alpha(\bar{\pi}) = \min_\pi \alpha(\pi)$$

se llama criterio *minimax.*

La siguiente afirmación es el análogo completo del teorema 2.11.2.

Teorema 2. *Supongamos que existe el criterio bayesiano* $\bar{\pi}$ *(correspondiente a cierta distribución a priori* $\bar{\mathbf{Q}}$*) para el cual*

$$\alpha_1(\bar{\pi}) = \ldots = \alpha_r(\bar{\pi}). \tag{10}$$

Entonces $\bar{\pi}$ *es el criterio minimax.*

Demostración. Designemos por $\bar{q}(j)$ las distribuciones a priori correspondientes a $\bar{\mathbf{Q}}$. Entonces para cualquier criterio π tenemos

$$\alpha(\pi) \geqslant \sum_{j=1}^{r} \bar{q}(j)\alpha_j(\pi) \geqslant \sum_{j=1}^{r} \bar{q}(j)\alpha_j(\bar{\pi}) = \max_j \alpha_j(\bar{\pi}) = \alpha(\bar{\pi}). \quad \triangleleft$$

La distribución $\bar{\mathbf{Q}} = \{\bar{q}(j)\}$ correspondiente al criterio $\bar{\pi}$ se llama criterio *peor* (o criterio *menos favorable*, compárese con el § 2.11). Esto está relacionado con el hecho de que para $\mathbf{Q} = \bar{\mathbf{Q}}$ se alcanza

$$\max_Q \alpha_Q(\pi^Q) = \max_Q \min_\pi \alpha_Q(\pi),$$

así que el criterio minimax (10) es el criterio bayesiano que posee la mayor probabilidad de equivocarse. La demostración de este hecho se puede hallar en los capítulos posteriores, donde también mostraremos que la peor distribución y el criterio minimax siempre existen.

Sin embargo, es preciso señalar que a distinción de los criterios bayesianos, los criterios minimax no randomizados existen no siempre, ni mucho menos. El asunto consiste en que las fronteras separadoras Γ_k de

los conjuntos $\tilde{\Omega}_k^Q$ (véase (6)) pueden tener una probabilidad no nula $\mathbf{P}_k(X \in \Gamma_k) > 0$ y, por lo tanto, los valores de $\alpha_k(\delta_Q)$, al modificarse continuamente $\mathbf{Q}$, pueden variar a saltos. Esto quiere decir, a su vez, que $r - 1$ ecuaciones $\alpha_1(\delta_Q) = \ldots = \alpha_r(\delta_Q)$ para $r - 1$ desconocidas $q(1), \ldots$ $\ldots, q(r-1)$ $\left(q(r) = 1 - \sum_{j=1}^{r-1} q(j)\right)$ pueden no tener solución. No obstante, en la clase de criterios bayesianos randomizados, el criterio minimax existe siempre. En calidad de ilustración examinaremos detalladamente esta cuestión (para el caso $r = 2$) en el párrafo siguiente.

Así pues, hemos hallado la forma explícita de los criterios bayesianos y hemos establecido que con su ayuda se pueden construir los criterios minimax. Resulta que de manera análoga también se pueden construir los criterios más potentes en las clases $K_{\alpha_1, \ldots, \alpha_{r-1}}$ introducidas en el punto 1.

4. Criterios más potentes. La definición del c.m.p. no randomizado fue dada en el punto 1. Aquí será cómodo extender esta definición a la clase de criterios randomizados. Supongamos que, análogamente al punto 1, $K_{\alpha_1, \ldots, \alpha_{r-1}}$ significa la clase de criterios *randomizados* con valores registrados de las probabilidades de los errores de j-ésimo género, $j = 1, \ldots, r - 1$:

$$K_{\alpha_1, \ldots, \alpha_{r-1}} = \{\pi\colon \alpha_j(\pi) = \alpha_j;\ j = 1, \ \ldots, \ r - 1\}.$$

Definición 8. El criterio $\pi_0 \in K_{\alpha_1, \ldots, \alpha_{r-1}}$ se llama *c.m.p. en* $K_{\alpha_1, \ldots, \alpha_{r-1}}$ si para cualquier $\pi \in K_{\alpha_1, \ldots, \alpha_{r-1}}$

$$\alpha_r(\pi_0) \leqslant \alpha_r(\pi).$$

Teorema 3. *Supongamos que existe una distribución* $\mathbf{Q} = \{q(1), \ldots$ $\ldots, q(r)\}$, *tal, que*

$$\alpha_j(\pi^Q) \equiv 1 - \mathbf{M}_j \pi_j^Q(X) = \alpha_j, \quad j = 1, \ \ldots, \ r - 1 \tag{11}$$

(en realidad, aquí tenemos $r - 1$ ecuaciones para los valores desconocidos de $q(1), \ \ldots, \ q(r-1)$). *Entonces el criterio bayesiano* π^Q, *definido en* (6) *y* (7), *será el más potente en la clase* $K_{\alpha_1, \ldots, \alpha_{r-1}}$.

Demostración. Según la definición del criterio bayesiano,

$$\alpha_Q(\pi^Q) \leqslant \alpha_Q(\pi).$$

Esto significa que para $\pi \in K_{\alpha_1, \ldots, \alpha_{r-1}}$ tendremos

$$\sum_{j=1}^{r} q(j)\alpha_j(\pi^Q) \leqslant \sum_{j=1}^{r-1} q(j)\alpha_j + q(r)\alpha_r(\pi).$$

Pero $\alpha_j(\pi^Q) = \alpha_j$ para $j \leqslant r - 1$ y, por consiguiente, $\alpha_r(\pi^Q) \leqslant \alpha_r(\pi)$. ◁

Aquí, por la misma causa que al hallar los criterios minimax, las ecuaciones (11) en la clase de los criterios no randomizados δ no siempre

son resolubles. En la clase de criterios randomizados, la situación cambia considerablemente. Esta circunstancia será ilustrada en el párrafo siguiente.

Ahora citemos el ejemplo de un problema real muy difundido, acerca de la verificación de un número finito de hipótesis simples.

Ejemplo 1. Supongamos que la hipótesis H_1 significa que un paciente que vino para ser reconocido por el médico, está sano, mientras que H_k significa que el paciente padece de cierta enfermedad A_k, $k \geqslant 2$. La tarea del médico consiste en aceptar una de las hipótesis H_j, basándose en las observaciones (que pueden ser escritas en forma del vector $\mathbf{x}_1 = (\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1s})$ que es de por sí la muestra multidimensional X de volumen unitario). Fijaremos las enfermedades A_k para que las hipótesis H_k sean simples y asimismo determinen por completo la distribución de la muestra X. Si el médico acepta la hipótesis H_k, $k \geqslant 2$, mientras que en realidad es cierta la hipótesis H_1, entonces cometerá un error de un tipo. Pero si, al contrario, reconoce sano (H_1) al enfermo (H_k), entonces cometerá un error de otro género. No es difícil comprender que los "efectos" producidos por los errores de estos dos tipos pueden ser muy diferentes.

De los resultados expuestos anteriormente deducimos que para construir la mejor regla de decisión, debemos saber las distribuciones del vector de las magnitudes observables $(\mathbf{x}_{11}, \ldots, \mathbf{x}_{1s})$ para individuos sanos y para individuos que padecen de la enfermedad A_k (para ello necesitamos muchos datos estadísticos de exámenes médicos). Por supuesto que una gran parte del problema aquí consiste en la propia elección de s y de las observaciones $(\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1s})$. Precisamente en esto se manifiesta principalmente el arte y la experiencia de los médicos.

Si el vector $(\mathbf{x}_{11}, \ldots, \mathbf{x}_{1s})$ se ha elegido de manera bastante argumentada, los teoremas 1—3 nos indicarán la vía directa para algoritmizar los problemas de la diagnosis de las enfermedades.

§ 2. Verificación de las hipótesis simples

En este párrafo examinaremos un poco más detalladamente un caso particular, cuando se verifican $r = 2$ hipótesis simples.

En los problemas de verificación de las hipótesis, estas últimas desempeñan a menudo un papel asimétrico, como ocurrió, digamos, en el ejemplo 1.1. Por eso, una de las hipótesis, por ejemplo H_1, suele llamarse *fundamental* y las demás, *alternativas*. En este caso, la probabilidad del error de primer género $\alpha_1(\delta)$ del criterio δ también se denomina *dimensión*, y el número $1 - \alpha_1(\delta)$, *nivel* del criterio. El número $\beta(\delta) = 1 - \alpha_2(\delta)$ se llama *potencia* del criterio.

La región $\Omega_2 \subset \mathscr{X}^n$ de aceptación de la hipótesis H_2 por el criterio no randomizado δ, en el caso de $r = 2$ se denomina *región crítica*. La probabili-

dad $\mathbf{P}_2(X \in \Omega_2)$ de caer en esta región, cuando es cierta H_2, equivale a la potencia del criterio $\beta(\delta)$. De aquí procede la denominación de "criterio más potente" para el criterio δ con el que $\beta(\delta)$ alcanza su máximo para un nivel registrado del criterio δ.

Señalemos ahora, que en el caso de $r = 2$, cualquier criterio, incluso el randomizado, puede caracterizarse por una función numérica. En efecto, el criterio randomizado arbitrario $\pi(x)$ se define totalmente por el valor de su r coordenadas $\pi_1(x), \ldots, \pi_r(x))$. Pero como $\Sigma\pi_j(x) = 1$, en caso de $r = 2$ es suficiente designar una función, digamos, $\pi_2(x)$. Esta función determina la probabilidad de que se acepte la alternativa H_2. Designémosla por $\pi(x)$ y llamémosla *función crítica* del criterio π que designaremos con la misma letra π. Es evidente que para los criterios no randomizados, $\pi(x)$ sólo adopta los valores de 0 y 1; en el caso general $0 \leqslant \pi(x) \leqslant 1$.

La dimensión $\alpha_1(\pi)$ del criterio π (o δ) y su potencia $\beta(\pi)$ se expresan a través de $\pi(x)$ del modo siguiente:

$$\alpha_1(\pi) = \mathbf{M}_1\pi(X), \quad \beta(\pi) = 1 - \alpha_2(\pi) = \mathbf{M}_2\pi(X).$$

Designemos por Z la relación de verosimilitud

$$Z = Z(x) = f_2(x)/f_1(x)$$

que examinaremos sólo para los valores de x, con los cuales ella está definida, o sea, para x cuando $f_1(x) + f_2(x) > 0$.

Teorema 1. 1) *Supongamos que* $c = q(1)/q(2)$, *donde* $\mathbf{Q} = (q(1), q(2))$, *y que* $q(2) = 1 - q(1)$ *es una distribución a priori dada. Entonces el criterio* $\pi_{c,p}$ *con la función crítica*

$$\pi_{c,p}(x) = \begin{cases} 1, & si \quad Z(x) > c, \\ p(x), & si \quad Z(x) = c, \\ 0, & si \quad Z(x) < c, \end{cases} \tag{1}$$

para cualquier función medible $p(x)$, $0 \leqslant p(x) \leqslant 1$, *es bayesiano para la distribución* $\mathbf{Q}$: $\pi_{c,p} = \pi^{\mathbf{Q}}$.

Los parámetros $\alpha_1(\pi_{c,p})$ y $\alpha_2(\pi_{c,p})$ *del criterio* $\pi_{c,p}$ *satisfacen la desigualdad*

$$\sum_{j=1}^{2} q(j)\alpha_j(\pi_{c,p}) \leqslant \sum_{j=1}^{2} q(j)(1 - g_j) \tag{2}$$

para todos $g_j \geqslant 0$, $g_1 + g_2 = 1$.

2) *Para* $\varepsilon > 0$ *dado, tal que* $\mathbf{P}_1(Z > 0) \geqslant \varepsilon$, *existen* $c > 0$ *y* $p(x) \equiv p = \text{const}$ *tales que* $\pi_{c,p} \in K_\varepsilon = \{\pi: \alpha_1(\pi) = \varepsilon\}$, *y* $\pi_{c,p}$ *es el c.m.p. en* K_ε. *Los números c y p se definen como la solución de la ecuación*

$$\alpha_1(\pi_{c,p}) \equiv \mathbf{M}_1\pi_{c,p}(X) \equiv \mathbf{P}_1(Z(X) > c) + p\mathbf{P}_1(Z(X) = c) = \varepsilon. \tag{3}$$

En este caso la potencia del criterio $\beta(\pi_{c,p}) = 1 - \alpha_2(\pi_{c,p})$ *satisface la desigualdad*

$$\beta(\pi_{c,p}) \geqslant \varepsilon. \tag{4}$$

Si no se cumple la relación $f_2(x) = f_1(x)$ c.d. $[\mu]$, *entonces, las desigualdades* (4) *y* (2) *para* $0 < q_1 < 1$ *son estrictas.*

El criterio $\pi_{c,p}$ *minimiza la probabilidad del error de primer género* $\alpha_1(\pi)$ *en la clase* K *de todos los criterios* π *con una probabilidad fija del error de segundo género:* $K = \{\pi: \alpha_2(\pi) = \alpha_2(\pi_{c,p})\}$.

3) *Existen* $c > 0$ *y* $p(x) \equiv p =$ const *tales, que el criterio* $\pi_{c,p}$ *será minimax. Los números* c *y* p *se determinan de la ecuación* $\alpha_1(\pi_{c,p}) = \alpha_2(\pi_{c,p})$ *o bien, que es lo mismo, de la ecuación*

$$\mathbf{P}_1(Z(X) > c) + \mathbf{P}_2(Z(X) > c) + p[\mathbf{P}_1(Z(X) = c) + \mathbf{P}_2(Z(X) = c)] = 1. \tag{5}$$

Es evidente que si la $\mathbf{P}_1$-distribución de $Z(X)$ es continua, o sea, si $\mathbf{P}_1(Z(X) = c) = 0$ para todos $c \geqslant 0$, entonces, en las dos últimas afirmaciones del teorema podemos poner $p \equiv 1$ ó $p \equiv 0$.

Nótese también que

$$\mathbf{P}_1(Z(X) = c) = \\ = \int\limits_{Z(x)=c} f_1(x)\mu^n(dx) = \int\limits_{Z(x)=c} \frac{f_2(x)}{c}\mu^n(dx) = \frac{1}{c}\mathbf{P}_2(Z(X) = c),$$

así que la continuidad en $(0, \infty)$ de la $\mathbf{P}_1$-distribución de Z conduce a la continuidad de la $\mathbf{P}_2$-distribución de Z.

El criterio $\pi_{c,p}$, basado en la relación de verosimilitud Z, se llama *criterio de la relación de verosimilitud.*

El teorema 1 muestra que todos los *criterios óptimos son criterios de la relación de verosimilitud.*

La segunda afirmación del teorema 1 lleva el nombre de *lema de Neyman — Pearson.* Si en esta afirmación, la condición $\mathbf{P}_1(Z > 0) \geqslant \varepsilon$ no se cumple, o sea, si $\mathbf{P}_1(Z = 0) = 1 - \delta$, $\delta < \varepsilon$, entonces el c.m.p. $\pi(x) = I_{\{Z(x)>0\}}$ tendrá potencia 1 y dimensión $\delta < \varepsilon$. Por ejemplo, si los portadores de las distribuciones $\mathbf{P}_1$ y $\mathbf{P}_2$ son disjuntos, entonces $Z = 0$ en el conjunto donde $f_1(x) > 0$ y, por lo tanto, $\mathbf{P}_1(Z > 0) = 0$. En este caso, las hipótesis H_1 y H_2 se distinguen por una observación, con probabilidades de errores iguales a cero, o sea, se distinguen de un modo determinado.

Demostración del teorema 1. La primera afirmación del teorema es el corolario directo del teorema 1.1A.

Para demostrar la segunda afirmación se puede hacer uso del teorema 1.3. Mostremos primeramente que la ecuación (3) es siempre resoluble respecto a c y p. Es evidente que la función $\varphi(c) = \mathbf{P}_1(Z > c)$ no crece en $[0, \infty)$. La variable aleatoria Z es propia con respecto a la distribución $\mathbf{P}_1$,

o sea,

$$\varphi(c) = \mathbf{P}_1(Z > c) =$$
$$= \int\limits_{Z(x)>c} f_1(x)\mu^n(dx) < \frac{1}{c} \int\limits_{Z(x)>c} f_2(x)\mu^n(dx) = \frac{1}{c}\mathbf{P}_2(Z > c) \to 0$$

cuando $c \to \infty$. Como, según la condición, $\varphi(0) \geqslant \varepsilon$, entonces existirá $c_\varepsilon \in \in (0, \infty)$ tal, que ($\varphi(c)$ será continua a la derecha)

$$\varphi(c_\varepsilon - 0) \geqslant \varepsilon, \quad \varphi(c_\varepsilon) \leqslant \varepsilon. \tag{6}$$

Si en (3) suponemos que $c = c_\varepsilon$, y designamos $\Delta_\varepsilon = \varphi(c_\varepsilon - 0) - \varphi(c_\varepsilon)$, obtendremos

$$\alpha_1(\pi_{c_\varepsilon,p}) = \varphi(c_\varepsilon) + p\Delta_\varepsilon.$$

Es evidente que aquí, en virtud de (6), siempre se puede escoger $p \in [0, 1]$ de modo*) que $\varphi(c_\varepsilon) + p\Delta_\varepsilon = \varepsilon$.

Ahora podemos proceder igualmente que en la demostración del teorema 1.3. Pongamos $q(1) = q_\varepsilon = c_\varepsilon/(c_\varepsilon + 1)$ y fijemos el p que hemos elegido. Entonces, el criterio $\pi_{c_\varepsilon,p}$ será bayesiano, correspondiente a la distribución $\mathbf{Q}_\varepsilon = (q_\varepsilon, 1 - q_\varepsilon)$ y al mismo tiempo $\alpha_1(\pi_{c_\varepsilon,p}) = \varepsilon$. Esto significa, en virtud del teorema 1.3, que $\pi_{c_\varepsilon,p}$ es el c.m.p. en K_ε.

Si tomamos el criterio $\pi(x) \equiv \varepsilon$, obtenemos

$$\pi \in K_\varepsilon, \ \alpha_2(\pi_{c_\varepsilon,p}) \leqslant \alpha_2(\pi) \equiv 1 - \varepsilon, \ \beta(\pi_{c_\varepsilon,p}) \geqslant \varepsilon.$$

No es otra cosa sino la desigualdad (2) ((1.8)) para $g_2 = \varepsilon$. Por consiguiente, si la relación $f_2(x) = f_1(x)$ c.d. $[\mu]$ no se cumple, entonces estas desigualdades serán estrictas. La afirmación del teorema acerca de la minimización de $\alpha_1(\pi)$ en el criterio $\pi_{c,p}$ de la clase $K = \{\pi\colon \alpha_2(\pi) = \alpha_2(\pi_{c,p})\}$ se deduce de los razonamientos anteriormente aducidos y de la simetría con respecto a las hipótesis H_1 y H_2 del planteamiento del problema en la primera afirmación del teorema.

A fin de demostrar la tercera afirmación del teorema 1 conviene valerse del teorema 1.2. Para esto sólo necesitamos comprobar si la ecuación $\alpha_1(\pi_{c,p}) = \alpha_2(\pi_{c,p})$ es resoluble respecto a c y p. Esta ecuación se puede escribir en la forma

$$\mathbf{M}_1\pi_{c,p}(X) = 1 - \mathbf{M}_2\pi_{c,p}(X)$$

o bien, que es lo mismo, en la forma de (5). Su solubilidad se deter-

---

*) Está claro que si $\varphi(c)$ es continua en el punto c_ε, el problema de resolución de (3) se reduce a la determinación de la cuantila de distribución de Z de orden $1 - \varepsilon$.

mina al igual que la solubilidad de la ecuación (3). Sólo es necesario señalar que siempre $\mathbf{P}_1(Z > 0) + \mathbf{P}_2(Z > 0) \geqslant 1$, ya que $\mathbf{P}_2(Z > 0) = = \int_{f_1(x)>0} f_2(x)\mu^n(dx) = 1$. ◁

Hemos visto una vez más que el objetivo de la introducción de los criterios bayesianos randomizados consiste en asegurar la variación "continua" de los parámetros de dichos criterios (los posibles valores de las dimensiones de los criterios $\pi_{c,p}$ llenan todo el intervalo $(0, 1)$). La falta de tal variación continua de los parámetros, relacionada con el hecho de que en el conjunto de la $\mathbf{P}_1$-probabilidad positiva es posible la igualdad $f_1(x) = = cf_2(x)$, constituye el principal obstáculo al hallar los criterios de un nivel dado o los minimax en la clase de criterios no randomizados. Este cuadro también se conserva por completo en el caso de verificación de un número mayor de hipótesis.

También es importante señalar que dos tipos de criterios óptimos — los más potentes y los minimax — resultan bayesianos en unas u otras distribucions a priori. Tampoco es difícil notar que la clase de todos los criterios más potentes coincide, desde cierto punto de vista, con la clase de todos los criterios bayesianos. Tal situación, en la que en calidad de base para la elección de los criterios óptimos puede utilizarse el enfoque bayesiano, también se conservará en mucho posteriormente.

Ejemplo 1. Examinemos el ejemplo 2 citado en la introducción. En este ejemplo, las hipótesis H_1 y H_2 tienen la forma $H_1 = \{x_i \in F(x)\}$, $H_2 = = \{x_i \in F(x - a)\}$, donde $F(x)$ es una función dada de distribución, y a, un número dado. Supongamos que $F(x)$ tiene densidad $f(x)$ y que la variable aleatoria $f(x_1 - a)/f(x_1)$ tiene una distribución continua. Entonces, según el lema de Neyman — Pearson (punto 2 del teorema 1), entre todos los criterios de nivel $1 - \varepsilon$, el criterio

$$\prod_{i=1}^{n} \frac{f(x_i - a)}{f(x_i)} \geqslant c_\varepsilon$$

será el más potente en el problema sujeto a examen, dedicado a la verificación de la hipótesis H_1 (falta el objeto), frente a la hipótesis H_2 (el objeto está presente). El número c_ε se determina de la condición

$$\mathbf{P}_1\left(\sum_{i=1}^{n} \ln \frac{f(x_i - a)}{f(x_i)} > \ln c_\varepsilon\right) = \varepsilon.$$

Si n son grandes, para el cálculo de esta probabilidad podemos, evidentemente, hacer uso del teorema central del límite.

§ 3*. Dos enfoques asintóticos del cálculo de los criterios. Comparación numérica

1. Observaciones preliminares. En los §§ 1 y 2 hemos hallado la forma de los criterios óptimos para verificar las hipótesis simples. El término "cálculo de los criterios" que hemos usado en el encabezamiento significará el cálculo de los parámetros que caracterizan el criterio. En el problema del c.m.p. esto es, en caso de $r = 2$, la búsqueda de las magnitudes c_ε y p para $\varepsilon > 0$ dado la determinación de la probabilidad del error de segundo género $\alpha_2(\pi_{c_\varepsilon,p})$ o bien, que es lo mismo, de la potencia del criterio $\beta(\pi_{c_\varepsilon,p}) = 1 - \alpha_2(\pi_{c_\varepsilon,p})$. La cuestión también puede ser planteada de una manera algo distinta. Hemos visto que en caso de $r = 2$ todos los criterios óptimos tienen la forma de las funciones $\pi_{c,p}$ representadas en (2.1). Supongamos que se da el criterio $\pi_{c,p}$. ¿Cómo determinar para él las probabilidades de los errores $\alpha_l(\pi_{c,p})$?

Esta misma pregunta también surge, por supuesto, en el caso general de $r > 2$ para el criterio (1.7), pero en este párrafo nos limitaremos, para abreviar, al caso de dos hipótesis simples.

Más abajo se examinan los enfoques asintóticos que permiten resolver *aproximadamente* (con grandes n) tales problemas. Esos mismos enfoques también pueden utilizarse para calcular los criterios que se examinarán en adelante.

Así pues, supongamos que se da el criterio (2.1) y que la distribución de $Z(X)$ es, para abreviar, continua, así que podemos poner $p = 1$. Entonces, el criterio (2.1) se volverá no randomizado (designémoslo por δ_c) y necesitaremos hallar sus valores:

$$\alpha_1(\delta_c) = \mathbf{P}_1\left(\frac{f_2(X)}{f_1(X)} \geqslant c\right), \tag{1}$$

$$\alpha_2(\delta_c) = \mathbf{P}_2\left(\frac{f_2(X)}{f_1(X)} < c\right).$$

Como $f_j(X) = \prod_{i=1}^{n} f_j(x_i)$, el suceso que se encuentra bajo el signo de probabilidad en (1) puede ser escrito en la forma

$$\sum_{i=1}^{n} \ln \frac{f_2(x_i)}{f_1(x_i)} \geqslant \ln c,$$

donde los sumandos

$$\eta_i = \ln \frac{f_2(x_i)}{f_1(x_i)}$$

son, evidentemente, variables aleatorias independientes, igualmente distribuidas en cada uno de los casos $X \in \mathbf{P}_j$, $j = 1, 2$.

Ahora bien, el hecho se reduce al estudio de las distribuciones de las sumas $\sum_{i=1}^{n} \eta_i$ de las variables aleatorias η_i.

En lo sucesivo supondremos que el volumen n de la muestra X crece indefinidamente. En esto caso, por criterio entendremos, en realidad, la sucesión de los criterios definidos para cada n (hemos utilizado ese mismo acuerdo para las estimaciones en el capítulo 2).

2. Hipótesis fijas. En este apartado supondremos que las distribuciones $\mathbf{P}_i$ están fijas, o sea, no dependen del volumen $n \to \infty$ de la muestra $X_n = [X_\infty]_n$. Examinemos el problema de cálculo del c.m.p. de nivel fijo $1 - \varepsilon$. Tenemos

$$\mathbf{M}_1\eta_i \equiv -a = \int f_1(x) \ln \frac{f_2(x)}{f_1(x)} \mu(dx) = -\varrho_1(\mathbf{P}_1, \mathbf{P}_2) < 0,$$

$$\mathbf{M}_2\eta_i \equiv b = \int f_2(x) \ln \frac{f_2(x)}{f_1(x)} \mu(dx) = \varrho_1(\mathbf{P}_2, \mathbf{P}_1) > 0,$$

donde ϱ_1 es la distancia de Kullback — Leibler (véase el § 2.21). Esto significa que, en virtud de la ley de los grandes números, la $\mathbf{P}_1$-distribución de $\frac{1}{n}\sum_{i=1}^{n} \eta_i$ permanecerá concentrada en el entorno del punto $-a$, y la $\mathbf{P}_2$-distribución, en el entorno del punto h. Y esta "separación" de las distribuciones será la mejor desde el punto de vista del lema de Neyman — Pearson. Designemos $\sigma_j^2 = \mathbf{D}_j\eta_1$ y supongamos que $\sigma_j^2 < \infty$. Entonces

$$\alpha_1(\delta_c) = \mathbf{P}_1 \left(\sum_{i=1}^{n} \eta_i \geqslant \ln c \right) = \mathbf{P}_1 \left(\frac{1}{\sigma_1\sqrt{n}} \sum_{i=1}^{n} (\eta_i + a) \geqslant \frac{\ln c + an}{\sigma_1\sqrt{n}} \right). \quad (2)$$

Escojamos en calidad de $c = c(n)$ toda sucesión para la cual

$$\frac{\ln c + an}{\sigma_1\sqrt{n}} \to \lambda_\varepsilon,$$

donde λ_ε es, como antes, la cuantila de la distribución normal de nivel $1 - \varepsilon$. Entonces, de (2) y del teorema central del límite resulta

$$\alpha_1(\delta_c) \sim 1 - \Phi\left(\frac{\ln c + an}{\sigma_1\sqrt{n}}\right) \to \varepsilon. \quad (3)$$

Definición 1. El criterio π que satisface la relación

$$\lim_{n\to\infty} \alpha_1(\pi) = \lim_{n\to\infty} \mathbf{M}_1\pi(X) = \varepsilon$$

se llama *criterio de nivel asintótico* $1 - \varepsilon$ (o de dimensión asintótica ε).

Por lo tanto, para

$$\ln c = -an + \lambda_\varepsilon \sigma_1 \sqrt{n} + o(\sqrt{n}), \tag{4}$$

el criterio δ_c tendrá el nivel asintótico $1 - \varepsilon$.

La relación (4) puede considerarse como la solución aproximada de la ecuación del número c_ε para el cual $\alpha_1(\delta_{c_\varepsilon}) = \varepsilon$.

Pongamos, para precisar, $\ln c = -an + \lambda_\varepsilon \sigma_1 \sqrt{n}$ y hallemos, para el c elegido, el comportamiento asintótico de la probabilidad del error de segundo género:

$$\alpha_2(\delta_c) = \mathbf{P}_2 \left(\sum_{i=1}^{n} \eta_i < \ln c \right) = \mathbf{P}_2 \left(\sum_{i=1}^{n} \eta_i < -an + \lambda_\varepsilon \sigma_1 \sqrt{n} \right) =$$

$$= \mathbf{P}_2 \left(\frac{1}{\sigma_2 \sqrt{n}} \sum_{i=1}^{n} (\eta_i - b) < (a + b)\sqrt{n}/\sigma_2 + \lambda_\varepsilon \sigma_1/\sigma_2 \right). \tag{5}$$

Como $-(a + b)\sqrt{n}/\sigma_2 + \lambda_\varepsilon \sigma_1/\sigma_2 \to -\infty$ cuando $n \to \infty$, aquí la aplicación del teorema central del límite sólo nos da que $\alpha_2(\delta_c) \to 0$.

El problema de cálculo del comportamiento asintóticamente exacto del segundo miembro en (5) conduce al problema de las probabilidades de grandes desviaciones para las sumas de variables aleatorias η_j.

Presentemos aquí los resultados de las probabilidades de grandes desviaciones, expuestos en el § 5 del capítulo 7 [11]. Supongamos que es necesario calcular el comportamiento asintótico $\mathbf{P}\left(\sum_{i=1}^{n} \xi_i > x \right)$ cuando $n \to \infty$, $x \to \infty$, donde ξ_i son independientes y están igual distribuidas. Admitamos que la distribución ξ_i tiene una componente absolutamente continua y que

$$\psi(\lambda) = \mathbf{M} e^{\lambda \xi} < \infty$$

para ciertos $\lambda > 0$. Supongamos, además, que

$$\lambda_+ = \sup \{\lambda \colon \psi(\lambda) < \infty\}, \tag{6}$$
$$\Lambda(\alpha) = -\inf_{\lambda} \{-\alpha\lambda + \ln \psi(\lambda)\},$$

y que $\lambda(\alpha)$ es el valor de λ con el que se alcanza este inf $\{\cdot\}$.

Entonces, es válida la afirmación siguiente. (Véanse los teoremas 9 y 10 del § 5 del capítulo 7 [11]. Las condiciones $\mathbf{D}\xi_i = 1$ y $\mathbf{M}\xi_i = 0$ que figuran en estos teoremas no desempeñan ningún papel).

Teorema 1. *Supongamos que* $\dfrac{x - n\mathbf{M}\xi_1}{\sqrt{n}} \to \infty$ *de modo que*

$$\limsup_{n \to \infty} \frac{x}{n} < \alpha_+ = \frac{\psi'(\lambda_+)}{\psi(\lambda_+)}.$$

Entonces la ecuación

$$\alpha\psi(\lambda) = \psi'(\lambda) \tag{7}$$

para el punto $\lambda(\alpha)$ *tiene, cuando* $\alpha < \alpha_+$, la única solución,

$$\mathbf{P}\left(\sum_{i=1}^{n}\xi_i > x\right) \sim \frac{1}{\sigma(\alpha)|\lambda(\alpha)|\sqrt{2\pi n}}\exp\{-n\Lambda(\alpha)\}, \tag{8}$$

donde

$$\alpha = \frac{x}{n},\ \sigma^2(\alpha) = \frac{\psi''(\lambda(\alpha))}{\psi(\lambda(\alpha))} - \alpha^2.$$

Además, son válidas las relaciones

$$\Lambda(\mathbf{M}\xi_1) = 0,\quad \Lambda'(\alpha) = \lambda(\alpha),$$

$$\Lambda''(\alpha) = \lambda'(\alpha) = \frac{\psi(\lambda(\alpha))}{\psi''(\lambda(\alpha)) - \alpha^2\psi(\lambda(\alpha))}.$$

Volvamos ahora al cálculo del comportamiento asintótico de la magnitud $\alpha_2(\delta_c)$ definida en (5) e igual a

$$\mathbf{P}_2\left(-\sum_{i=1}^{n}\eta_i > an - y\sqrt{n}\right) = \mathbf{P}_2\left(\sum_{i=1}^{n}(-\eta_i + b) > (a+b)n - y\sqrt{n}\right)$$

cuando $y = \lambda_\epsilon\sigma_1$. Para hacer uso del teorema expuesto es necesario poner

$$\xi_i = -\eta_i = \ln\frac{f_1(x_i)}{f_2(x_i)},\quad x = an - y\sqrt{n}.$$

Entonces, cuando $0 \leqslant \lambda \leqslant 1$, obtenemos

$$\psi(\lambda) = \mathbf{M}_2 e^{-\lambda\eta_1} = \int f_2(x)(f_1(x)/f_2(x))^\lambda\mu(dx) =$$

$$= \int f_1^\lambda(x)f_2^{1-\lambda}(x)\mu(dx) \leqslant \left(\int f_1(x)\mu(dx)\right)^\lambda\left(\int f_2(x)\mu(dx)\right)^{1-\lambda} = 1.$$

De aquí asimismo se deduce que $\psi(\lambda)$ también será finito en cierto entorno del punto $\lambda = 1$ si

$$\int f_1(x)(f_1(x)/f_2(x))^\gamma\mu(dx) < \infty \tag{9}$$

para cualquier $\gamma > 0$. Luego, la ecuación para el punto $\lambda(\alpha)$ tendrá la forma

$$-\alpha + \frac{\psi'(\lambda)}{\psi(\lambda)} = 0,$$

o bien

$$\psi'(\lambda) = \int f_2(x)(f_1(x)/f_2(x))^\lambda\ln\frac{f_1(x)}{f_2(x)}\mu(dx) =$$

$$= \alpha\int f_2(x)(f_1(x)/f_2(x))^\lambda\mu(dx). \tag{10}$$

Si $\alpha = a = \varrho_1(\mathbf{P}_1, \mathbf{P}_2) = \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} \mu(dx)$, entonces (10) será satisfecha cuando $\lambda = 1$. Esto quiere decir que

$$\lambda(a) = 1, \quad \psi(\lambda(a)) = \psi(1) = 1.$$

De aquí se desprende que

$$\Lambda(a) = a\lambda(a) - \ln \psi(\lambda(a)) = a,$$

$$\psi''(\lambda(a)) = \psi''(1) = \int f_1(x)\left(\ln \frac{f_1(x)}{f_2(x)}\right)^2 \mu(dx),$$

$$\sigma^2(a) = \psi''(1) - a^2 = \sigma_1^2,$$

$$\Lambda'(a) = \lambda(a) = 1, \ \Lambda''(a) = \sigma_1^{-2}.$$

Las condiciones del referido teorema se cumplirán si

1) la $\mathbf{P}_2$-distribución de $\ln \frac{f_1(x_1)}{f_2(x_1)}$ tiene una componente absolutamente continua,

2) $f_1(x)(f_1(x)/f_2(x))^\gamma \mu(dx) < \infty$ para cualquier $\gamma > 0$.

Teniendo en cuenta que en nuestro caso las funciones $\sigma(\alpha)$, $\lambda(\alpha)$, $\Lambda''(\alpha)$ son continuas en el entorno del punto $\alpha_2 = a$ y que $\alpha = x/n = a - y/\sqrt{n}$, obtenemos $\Lambda(\alpha) = a - \frac{y}{\sqrt{n}} + \frac{y}{2\sigma_1^2 n} + o\left(\frac{1}{n}\right)$.

Por lo tanto, ahora podemos enunciar el siguiente corolario del teorema citado.

Corolario 1'. *Supongamos que se cumple la condición* (9) *que la* $\mathbf{P}_2$*-distribución de* $\ln \frac{f_1(x_1)}{f_2(x_1)}$ *tiene una componente absolutamente continua. Entonces, cuando* $n \to \infty$,

$$\alpha_2(\delta_c) = \mathbf{P}_2\left(\sum_{i=1}^{n} \eta_i > an - y\sqrt{n}\right) \sim$$

$$\sim \frac{1}{\sigma_1\sqrt{2\pi n}} \exp\{-na + y\sqrt{n} - y^2/(2\sigma_1^2)\} =$$

$$= \frac{1}{\sigma_1\sqrt{2\pi n}} \exp\{-n\varrho_1(\mathbf{P}_1, \mathbf{P}_2) + \lambda_\varepsilon \sigma_1 \sqrt{n} - \lambda_\varepsilon^2/2\}. \qquad (11)$$

Ahora bien, $\alpha_2(\delta_c)$ decrece exponencialmente*) cuando $n \to \infty$.

No es difícil ver que si tomamos un c registrado en (1), ambas probabilidades $\alpha_1(\delta_c)$ y $\alpha_2(\delta_c)$ decrecerán exponencialmente, al igual que el valor de $\alpha_Q(\delta_Q)$ para cualquier Q registrado. Como

$$\mathbf{M}_1 e^{\lambda\eta_1} = \int f_1(x)\left(\frac{f_2(x)}{f_1(x)}\right)^{\lambda}\mu(dx) = \psi(1-\lambda),$$
$$\min_{\lambda}\psi(\lambda) = \min_{\lambda}\psi(1-\lambda),$$

entonces $\alpha_1(\delta_c)$ y $\alpha_2(\delta_c)$ decrecerán con igual velocidad (su dependencia de n será la misma). Esto quiere decir que el criterio minimax corresponderá a cierto c registrado, cuyo valor aproximado se determina fácilmente resolviendo la ecuación $\alpha_1(\delta_c) = \alpha_2(\delta_c)$ y utilizando el análisis asintótico del segundo miembro (8) cuando $\alpha = c/n$, $n \to \infty$.

La aproximación exponencial (11) actúa bastante bien con grandes n siempre que la desviación normalizada

$$\frac{x + n\mathbf{M}_2\eta}{\sigma_2\sqrt{n}} = \frac{\sqrt{n}}{\sigma_2}\left(\varrho_1(\mathbf{P}_1,\ \mathbf{P}_2) + \varrho_1(\mathbf{P}_2,\ \mathbf{P}_1)\right) - \frac{\lambda_0\sigma_1}{\sigma_2} \tag{13}$$

también sea grande (véase la enunciación del teorema).

En los problemas aplicados, donde el número n está limitado por valores del orden de 100, esta condición se cumple rara vez y el valor de (13) a menudo resulta comparable con 1. Esto dificulta la utilización del referido enfoque del cálculo de $\alpha_2(\delta_c)$ y corresponde a la situación en que el valor de $\alpha_2(\delta_c)$, junto con $\alpha_1(\delta_c)$, no es muy pequeño (tiene una magnitud comparable, digamos, con 0,1). Al mismo tiempo, los valores de n del orden de 100 son completamente suficientes para la aplicación satisfactoria del teorema central del límite en la zona de "desviaciones normales".

---

*) A la vez hemos obtenido la posibilidad de dar una definición más de la distancia de Kullback — Leibler:

$$\varrho_1(\mathbf{P}_1,\ \mathbf{P}_2) = -\lim_{n\to\infty}\frac{1}{n}\ln\alpha_2(\delta_c) = -\lim_{n\to\infty}\frac{1}{n}\inf_{\delta\in K_\varepsilon}\ln\alpha_2(\delta).$$

Con arreglo a esto se puede señalar que ese mismo orden de pequeñez $\exp\{-n\varrho_1(\mathbf{P}_1,\ \mathbf{P}_2)\}$ es propio de la $\mathbf{P}_2$-probabilidad de que la función em pírica de distribución F_n^* vaya a parar al entorno de la función de distribución F_1 correspondiente a $\mathbf{P}_1$. Mejor dicho, si $\delta = \delta(n) \to 0$ bastante lentamente, entonces

$$-\lim_{n\to\infty}\frac{1}{n}\ln\mathbf{P}_2\left(\sup_x|F_n^*(x) - F_1(x)| < \delta\right) = \varrho_1(\mathbf{P}_1,\ \mathbf{P}_2) \tag{12}$$

(teorema de Sanov). Por consiguiente, la distancia $\varrho_1(\mathbf{P}_1,\ \mathbf{P}_2)$ tiene un sentido probabilístico profundo. Superando ciertas dificultades, el lector puede obtener (del teorema 6, § 2, capítulo V en [11]) la demostración de la relación (12).

Ahora bien, la cuestión que nos interesa consiste en saber cuando podemos usar las aproximaciones normales

$$\alpha_1(\delta_c) = \mathbf{P}_1\left(\sum_{i=1}^{n}\eta_i \geqslant \ln c\right) \approx 1 - \Phi\left(\frac{\ln c - n\mathbf{M}_1\eta_1}{\sigma_1\sqrt{n}}\right),$$

$$\alpha_2(\delta_c) = \mathbf{P}_2\left(\sum_{i=1}^{n}\eta_i < \ln c\right) \approx \Phi\left(\frac{\ln c - n\mathbf{M}_2\eta_1}{\sigma_2\sqrt{n}}\right) \tag{14}$$

a fin de calcular ambos valores de $\alpha_1(\delta_c)$ y $\alpha_2(\delta_c)$.

Para fundamentar las fórmulas (14) surge otro enfoque basado en la suposición de que las hipótesis H_1 y H_2 son próximas.

3. Hipótesis próximas. Aquí examinaremos la muestra X en el esquema de series y estimaremos que las distribuciones $\mathbf{P}_1$ y $\mathbf{P}_2$ dependen de n de modo que

$$\varrho_1(\mathbf{P}_1, \mathbf{P}_2) + \varrho_1(\mathbf{P}_2, \mathbf{P}_1) \to 0 \tag{15}$$

cuando $n \to \infty$, y la sucesión (13) converge hacia el límite positivo finito.

Para facilitar los razonamientos y hacerlos útiles en la exposición ulterior, aquí nos limitaremos al caso paramétrico cuando $X \in \mathbf{P}_\theta$.

$$H_1 = \{\theta = \theta_1\}, \quad H_2 = \{\theta = \theta_2\},$$

y la familia $\{\mathbf{P}_\theta\}$ satisface las condiciones de regularidad (RR) (véase el § 2.24).

Hagamos primeramente algunas observaciones no formales que explican la esencia de la cuestión. Examinamos las hipótesis próximas, o sea, supongamos que $\theta_2 = \theta_1 + \delta$, donde δ es pequeño. En este caso, el logaritmo de la relación de verosimilitud, a base del cual se construye el c.m.p., puede representarse en la forma*)

$$\ln\frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} \sim \delta L'(X, \theta_1). \tag{16}$$

La estadística $U = L'(X, \theta_1)$, es decir, la parte principal en (16), es llama, a veces, *aporte eficiente*. Si la hipótesis H_1 es cierta, entonces

$$\mathbf{M}_{\theta_1}U = 0, \quad \mathbf{D}_{\theta_1}U = nI(\theta_1).$$

Como

$$L'(X, \theta_1) - L'(X, \theta_2) \sim \delta L''(X, \theta_2), \quad \mathbf{M}_{\theta_2}L''(X, \theta_2) = -nI(\theta_2),$$

entonces

$$\mathbf{M}_{\theta_2}U \sim \delta nI(\theta_2) \sim \delta nI(\theta_1),$$
$$\mathbf{D}_{\theta_2}U \sim nI(\theta_2) \sim nI(\theta_1).$$

*) El signo $\sim$, aquí utilizado, significa la equivalencia asintótica cuando $\delta \to 0$.

Esto quiere decir que las distribuciones de U para las hipótesis H_1 y H_2 y para grandes n serán distinguibles siempre que la magnitud $\mathbf{M}_{\theta_2}U - \mathbf{M}_{\theta_1}U \sim \delta n I(\theta_1)$ sea mucho mayor que $\sqrt{\mathbf{D}_{\theta_1}U} \sim \sqrt{nI(\theta_1)}$ o comparable con ésta. En otros términos, debe cumplirse la igualdad $\delta n = v\sqrt{n}$, $v \neq 0$, o bien, que es lo mismo, $\delta = v/\sqrt{n}$.

Así pues, pasando a una exposición más exacta, supongamos que

$$\theta_2 = \theta_1 + v/\sqrt{n}, \tag{17}$$

donde consideraremos registradas las magnitudes θ_1 y v.

Siguiendo las designaciones del capítulo 2, pongamos

$$Z_i(t) = \frac{f_{\theta_i+t}(X)}{f_{\theta_i}(X)}, \quad Y_i(v) = \ln Z_i\left(\frac{v}{\sqrt{n}}\right).$$

Entonces

$$\sum_{i=1}^{n} \eta_i = \ln \frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} = Y_1(v) = -Y_2(-v). \tag{18}$$

En virtud del teorema 2.29.3 para $X \in \mathbf{P}_{\theta_1}$, tenemos

$$Y_1(v) = \xi_n v - \frac{1}{2} v^2 (I(\theta_1) + \varepsilon_n), \tag{19}$$

donde $\varepsilon_n \underset{\mathbf{P}_{\theta_1}}{\to} 0$, $\xi_n I^{-1/2}(\theta_1) \in \Phi_{0,1}$. Análogamente, para $X \in \mathbf{P}_{\theta_2}$,

$$-Y_2(-v) = \xi_n v + \frac{1}{2} v^2 (I(\theta_2) + \varepsilon_n),$$

donde $\varepsilon_n \underset{\mathbf{P}_{\theta_2}}{\to} 0$, $\xi_n I^{-1/2}(\theta_2) \in \Phi_{0,1}$.

Como $I(\theta_2) \to I(\theta_1)$, obtenemos que para la hipótesis H_j, $j = 1, 2$,

$$\sum_{i=1}^{n} \eta_i \Rightarrow \xi |v| \sqrt{I(\theta_1)} + (-1)^j \frac{v^2}{2} I(\theta_1), \quad \xi \in \Phi_{0,1}.$$

Esto significa que del teorema 2.29.3 se deduce el

Corolario 2. *Supongamos que se cumplen las condiciones* (*RR*), (17). *Entonces, para cualquier c registrado son válidas las fórmulas* (14) *o bien, más exactamente,*

$$\begin{aligned} \alpha_1(\delta_c) &= \mathbf{P}_{\theta_1}\left(\sum_{i=1}^{n} \eta_i \geqslant \ln c\right) \to 1 - \Phi\left(\frac{\frac{v^2}{2} I(\theta_1) + \ln c}{|v|\sqrt{I(\theta_1)}}\right), \\ \alpha_2(\delta_c) &= \mathbf{P}_{\theta_2}\left(\sum_{i=1}^{n} \eta_i < \ln c\right) \to \Phi\left(\frac{-\frac{v^2}{2} I(\theta_1) + \ln c}{|v|\sqrt{I(\theta_1)}}\right). \end{aligned} \tag{20}$$

Definición 2. Los criterios π_1 y π_2 se llaman *equivalentes asintóticamente* si

$$\lim_{n\to\infty} \sup |\alpha_j(\pi_1) - \alpha_j(\pi_2)| = 0, \quad j = 1, 2.$$

El criterio π se llama *criterio asintóticamente más potente* (*c.a.m.p.*) si el mismo es asintóticamente equivalente al c.m.p.

En vista de que en las representaciones (18) y (19), $\xi_n = L'(X, \theta_1)n^{-1/2}$, de éstas se deduce que el criterio δ, con la región crítica

$$\frac{vL'(X, \theta_1)}{\sqrt{nI(\theta_1)}} > vd, \quad d = \frac{v^2 I(\theta) + 2\ln c}{2|v|\sqrt{I(\theta_1)}},$$

(aquí tiene importancia el signo de v) tendrá los mismos valores límites $\alpha_i(\delta)$ que el criterio δ_c y por consiguiente, será el c.a.m.p.

Además, en virtud de los resultados del § 2.29,

$$\xi_n = L'(X, \theta_1)/\sqrt{n} = (\hat{\theta}^* - \theta_1)\sqrt{nI(\theta_1)}(1 + \varepsilon_n(X, \theta_1)),$$

$\varepsilon_n(X, \theta_1) \underset{P_{\theta_1}}{\to} 0$. De aquí resulta que el criterio con la región crítica

$$v(\hat{\theta}^* - \theta_1)\sqrt{nI(\theta_1)} > vd, \tag{21}$$

también será el c.a.m.p.

Para obtener el c.m.p. δ_c de nivel asintótico $1 - \varepsilon$, es suficiente en (20) poner $d = \lambda_\varepsilon$. La probabilidad del error de segundo género $\alpha_2(\delta_c)$ convergerá hacia $\Phi(-v\sqrt{I(\theta_1)} + \lambda_\varepsilon)$.

Para $c = 1$ ambos límites en (20) tendrán el mismo valor:

$$\lim_{n\to\infty} \alpha_j(\delta_c) = \Phi(-v\sqrt{I(\theta_1)}/2).$$

En este caso, el criterio δ_c (compárese con el teorema 1.2) es natural llamarlo *asintóticamente minimax*.

4. Comparación de los enfoques asintóticos. Ejemplo numérico. En los apartados 2 y 3 hemos examinado dos enfoques asintóticos (cada uno de los cuales está justificado en determinadas condiciones) que permiten indicar los valores aproximados de las probabilidades de los errores de primero y segundo género del c.m.p.*) En el caso de *hipótesis registradas*, estas fórmulas se dan en (3) y (11), y en el caso de *hipótesis próximas*, en (14) y (20). Las fórmulas (11) y (20) son una aproximación secundaria en compa-

---

*) Nótese que a la par con los dos enfoques propuestos se puede examinar un espectro entero de casos intermedios, los cuales en el lenguaje paramétrico pueden representarse en la forma (compárese con (17)) $\theta_2 = \theta_1 + zn^{-\gamma}$, $0 \leq \gamma \leq 1/2$. Las hipótesis próximas de tal género representan interés al seleccionar las fórmulas aproximadas que reflejan lo más exactamente una situación concreta dada.

ración con (8) y (14), por eso es necesario, en la medida de lo posible, dar preferencia a estas últimas.

Ya hemos señalado que para pequeños valores de $\alpha_1(\delta)$, $\alpha_2(\delta)$ (digamos, del orden de 0,01 y menos) conviene más utilizar el enfoque relacionado con las hipótesis registradas. Aquí es importante tener una precisión relativa de aproximación bastante buena, la cual es asegurada por las fórmulas (8) y no es garantizada por el teorema central del límite. No obstante, si $\alpha_1(\delta)$ y $\alpha_2(\delta)$ son comparables con 0,1 (digamos, $\geqslant 0{,}1$), se puede recomendar el segundo enfoque, considerando la segunda hipótesis dada $H_2 = \{\theta = \theta_2\}$ como un elemento de la sucesión de las hipótesis próximas $H_{2,n} = \{\theta = \theta_1 + v/\sqrt{n}\}$, donde, evidentemente, es necesario, para θ_1 y θ_2 dados, poner $v = \sqrt{n}(\theta_2 - \theta_1)$. Como los valores $\alpha_1(\delta)$ y $\alpha_2(\delta)$ esperados no son muy pequeños, el valor absoluto de $v/\sqrt{I(\theta_1)}$ no debe ser grande.

Ejemplo 1. Citemos ahora un ejemplo numérico que ilustra, en cierta medida, la relación existente entre los dos métodos de aproximación propuestos anteriormente.

Supongamos que $X \in \Gamma_{\theta,1}$, o sea, que x_i tienen una densidad

$$f_\theta(x) = \theta e^{-\theta x}, \quad x \geqslant 0,$$

y la hipótesis fundamental H_1 tienen la forma $H_1 = \{\theta = 1\}$. En calidad de alternativas examinemos las hipótesis simples $H_2^{(1)} = \{\theta = 0{,}5\}$, $H_2^{(2)} = \{\theta = 0{,}8\}$, $H_2^{(3)} = \{\theta = 0{,}9\}$.

Basándose en la muestra X, la hipótesis H_1 se verificará frente a una de las hipótesis $H_2^{(j)}$, $j = 1, 2, 3$. Ahora bien, aquí $\theta_1 = 1$, y para θ_2 hay tres variantes: $\theta_2 = 0{,}5$, $\theta_2 = 0{,}8$ y $\theta_2 = 0{,}9$, las dos últimas de las cuales trataremos de examinarlas como correspondientes a las hipótesis "próximas" a H_1. Realicemos el cálculo de los criterios para las muestras de volúmenes $n = 30, 100, 300, 1000$.

En nuestro caso

$$\eta_i = \ln \frac{f_{\theta_2}(x_i)}{f_{\theta_1}(x_i)} = \ln \theta_2 - (\theta_2 - 1)x_i, \tag{22}$$

$$l'(\theta_1, x_i) = 1 - x_i, \tag{23}$$

$$\theta^* = 1/\bar{x}.$$

De aquí resulta que el c.m.p. δ_c, así como ambos c.a.m.p. examinados anteriormente (con regiones críticas en forma de

$$\sum l'(x_i, \theta_1) < d_1 \quad \text{y} \quad \theta^* - \theta_1 < d_1/(nI(\theta_1)), \ d_1 = d\sqrt{nI(\theta_1)}),$$

tendrán el aspecto: $\delta_c(X) = H_2^{(j)}$ si

$$\sum_{i=1}^{n} (x_i - 1) > d_1. \tag{24}$$

Si $X \in \Gamma_{1,1}$ (hipótesis H_1), entonces

$$\mathbf{M}_1 x_1 = 1, \quad \mathbf{D}_1 x_1 = 1 = I(1) = \mathbf{M}_1 [l'(x_1, 1)]^2.$$

Por lo tanto, si ponemos $d_1 = 2\sqrt{n}$, entonces (compárese con (14))

$$\alpha_1(\delta_c) = \mathbf{P}_1 \left(\sum_{i=1}^{n} (x_i - 1) > d_1 \right) =$$

$$= \mathbf{P}_1 \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_i - 1) > 2 \right) \to 1 - \Phi(2) \approx 0{,}023 \qquad (25)$$

cuando $n \to \infty$. Como en nuestro caso $\sum_{i=1}^{n} \eta_i = n \ln \theta_2 + (1 - \theta_2) \sum_{i=1}^{n} x_i$, entonces $\ln c$ en (14) (o en (20)) está ligado a d_1 mediante la relación

$$\ln c = n(\ln \theta_2 + 1 - \theta_2) + (1 - \theta_2) d_1.$$

A continuación presentamos tres tablas. En todas d_1 se supone elegido de modo que se cumple (25) (o sea, $d_1 = 2\sqrt{n}$). En la primera tabla se comparan los valores verdaderos de $\alpha_1(\delta_c)$ con la aproximación (25). En la segunda tabla se dan los valores verdaderos de la probabilidad del error de segundo género $\alpha_2(\delta_c)$ y de la aproximación para $\alpha_2(\delta_c)$, obtenidos por las fórmulas de las grandes desviaciones (8). En la tercera tabla se comparan los valores verdaderos de $\alpha_2(\delta_c)$ con las aproximaciones obtenidas por las fórmulas de las hipótesis próximas (14). Nótese que aquí utilizamos las aproximaciones (8) y (14) sin hacer uso de las aproximaciones secundarias (11) y (20) que contienen errores adicionales. Todos los cálculos necesarios se exponen más adelante.

Los números en las tablas 1—3 se dan con una exactitud de hasta dos cifras significativas después de la coma.

Tabla 1. Valores de $\alpha_1(\delta_c)$. Renglón superior: valores verdaderos; renglón inferior: valores aproximaciones (14)

n	30	100	300	1000
	0,031 0,023	0,028 0,023	0,026 0,023	0,024 0,023

Tabla 2. Valores de $\alpha_2(\delta_c)$. Renglón superior: valores verdaderos; renglón inferior: valores aproximaciones (8) o (26) (grandes desviaciones)

θ_2 \ n	30	100	300	1000
0,5	0,028 0,033	$15\cdot10^{-7}$ $15\cdot10^{-7}$	$19\cdot10^{-19}$ $19\cdot10^{-19}$	$18\cdot10^{-72}$ $18\cdot10^{-72}$
0,8 —	0,71 —	0,35 —	0,028 0,033	$33\cdot10^{-8}$ $34\cdot10^{-8}$
0,9	0,89 —	0,79 —	0,53 —	0,085 0,11

La comparación de las tablas 2 y 3 muestra que de acuerdo con las observaciones hechas anteriormente, la aproximación basada en grandes desviaciones actúa mejor en la parte derecha superior de la tabla (donde $(\theta_1 - \theta_2)\sqrt{n} = (1 - \theta_2)\sqrt{n} > 3$), mientras que la aproximación basada en hipótesis próximas actúa mejor en la parte izquierda inferior de la tabla (donde $(1 - \theta_2)\sqrt{n} < 3$). Las rayas en las tablas están puestas allí donde la aplicación del referido enfoque no tiene sentido (en la tabla 2, por ejemplo, la aproximación (8) no se aplica en todos los casos cuando $\alpha_2(\delta_c) > 0{,}1$). El cálculo de $\alpha_2(\delta_c)$, cuando este valor es, digamos, menor de 10^{-6}, rara vez tiene sentido práctico. En la tabla 2 hemos calculado valores muy pequeños de $\alpha_2(\delta_c)$, cuando $\theta_2 = 0{,}5$, $n = 300$, 1000, únicamente a fin de comparar los resultados de los cálculos.

Tabla 3. Valores de $\alpha_2(\delta_c)$. Renglón superior: valores verdaderos; renglón inferior: valores aproximaciones (14) (hipótesis semejantes)

θ_2 \ n	30	100	300	1000
	0,028	$15\cdot10^{-7}$	$19\cdot10^{-19}$	$18\cdot10^{-72}$
0,5	0,041	$31\cdot10^{-6}$	—	—
	0,71	0,35	0,028	$33\cdot10^{-8}$
0,8	0,69	0,35	0,031	$12\cdot10^{-7}$
	0,89	0,79	0,53	0,085
0,9	0,89	0,79	0,52	0,086

Para acabar con los comentarios dedicados a las tablas, es preciso explicar cómo hemos calculado los valores verdaderos $\alpha_i(\delta_c)$, $i = 1, 2$ y en qué se transforman las aproximaciones (8) y (14) en nuestro caso concreto.

El valor de $\alpha_2(\delta_c)$ es igual a

$$\alpha_2(\delta_c) = \mathbf{P}_{\theta_2}\left(\sum_{i=1}^{n}(x_i - 1) < 2\sqrt{n}\right).$$

Como $\mathbf{M}_{\theta_2}x_i = 1/\theta_2 \mathbf{D}_{\theta_2}x_i = 1/\theta_2^2$, la aproximación normal (14) para $\alpha_2(\delta_c)$ tiene la forma

$$\Phi\left(\frac{\theta_2}{\sqrt{n}}\left[\left(1 - \frac{1}{\theta_2}\right)n + 2\sqrt{n}\right]\right) = \Phi((\theta_2 - 1)\sqrt{n} + 2\theta_2).$$

Examinemos ahora la fórmula (8) en la que en nuestro caso es necesario poner $\xi_i = x_i$, $x = -n - 2\sqrt{n}$. Aquí, la condición del teorema 1,

$$\frac{x - n\mathbf{M}\xi_1}{\sqrt{n}} = \frac{-n - 2\sqrt{n} + n/\theta_2}{\sqrt{n}} = \sqrt{n}\left(\frac{1 - \theta_2}{\theta_2}\right) - 2 \to \infty,$$

se cumple. Seguidamente,

$$\psi(\lambda) = \mathbf{M}_{\theta_2} e^{-\lambda x_i} = \theta_2 \int_0^\infty e^{-\lambda x - \theta_2 x} dx = \frac{\theta_2}{\lambda + \theta_2},$$

$$\lambda_+ = \infty, \quad \alpha_+ = \lim_{\lambda\to\infty} \frac{\psi'(\lambda)}{\psi(\lambda)} = 0,$$

$$\alpha = \frac{x}{n} = -1 - \frac{2}{\sqrt{n}}.$$

Como $\lim_{n\to\infty} \alpha = -1 < 0$, la condición $\lim_{n\to\infty} \sup \frac{x}{n} < \alpha_+$ también se cumple. En nuestro caso la ecuación (7) tiene la forma

$$\frac{\alpha\theta_2}{\lambda + \theta_2} = -\frac{\theta_2}{(\lambda + \theta_2)^2},$$

y su solución es $\lambda(\alpha) = -1/\alpha - \theta_2$. De aquí hallamos

$$\Lambda(\alpha) = -\ln(-\alpha\theta_2) - 1 - \alpha\theta_2, \quad \sigma^2(\alpha) = 1/\lambda'(\alpha) = \alpha^2.$$

Ahora bien, en virtud de (8) obtenemos

$$\alpha_2(\delta_c) = \mathbf{P}_{\theta_1}\left(\sum_{i=1}^n \xi_i > x\right) = \mathbf{P}_{\theta_2}\left(\sum_{i=1}^n (x_i - 1) < 2\sqrt{n}\right) \sim$$

$$\sim \frac{1}{(1 + \alpha\theta_2)\sqrt{2\pi n}} \exp\{n[\ln(-\alpha\theta_2) + 1 + \alpha\theta_2]\}. \qquad (26)$$

Suponiendo aquí $\alpha = -1 - 2\sqrt{n}$, obtenemos las fórmulas con las que hemos calculado los valores de $\alpha_2(\delta_c)$ en la tabla 2 (renglón inferior).

Señalemos, para comparar, que el segundo miembro de (11) en nuestro caso se transforma en la expresión

$$\frac{1}{(1 - \theta_2)\sqrt{2\pi n}} \exp\{n[\ln \theta_2 + 1 - \theta_2] + 2(1 - \theta_2)\sqrt{n} - 2\} \qquad (27)$$

que puede ser obtenida de (26), sustituyendo allí $\alpha = -1 - 2/\sqrt{n}$ y eliminando, después del desarrollo en serie, los términos del orden de $1\sqrt{n}$ y superiores.

En el denominador (26), el primer factor igual a $\sigma(\alpha)|\lambda(\alpha)| = 1 + \alpha\theta_2 = 1 - \theta_2 - 2\theta_2/\sqrt{n}$ se sustituye en (27) por $\sigma_1 = 1 - \theta_2$. Si θ_2 es próximo a 1, el error relativo, relacionado con el sumando de corrección $-2\theta_2/\sqrt{n}$, puede resultar considerable. Por ejemplo, para $\theta_2 = 0{,}8$, $n = 100$ obtenemos $2\theta_2/\sqrt{n} = 0{,}16$, $\sigma_1 = 1 - \theta_2 = 0{,}2$, $\sigma(\alpha)|\lambda(\alpha)| = 0{,}2 - 0{,}16 = 0{,}04$, así que el primer factor en (27) es 5 veces (!) mayor que en (26). Este ejemplo muestra que en el caso de hipótesis semejantes, cuando el factor σ_1 en (11) es pequeño, las aproximaciones (11) o (27) deben utilizarse con mucho cuidado.

Para calcular los valores verdaderos de $\alpha_i(\delta_c)$ hemos usado el hecho siguiente. Sea $\eta(t)$ el proceso de reconstrucción (véase [11]) para errar a saltos $x_1, x_2, \ldots$, o sea,

$$\eta(t) = \min\left\{k: \sum_{i=1}^k x_i \geqslant t\right\}.$$

En este caso, si $x_i \in \Gamma_{\theta,1}$, entonces, como hemos mostrado en el § 4 del capítulo 13

[11], el proceso $\xi(t) = \eta(t) - 1$ es, para $t > 0$, el proceso de Poisson con parámetro θ, o sea,

$$\mathbf{P}(\eta(t) - 1 = k) = e^{-\theta t}\frac{(\theta t)^k}{k!}.$$

Ahora señalemos que $\left\{\sum_{i=1}^{n} x_i \geq t\right\} = \{\eta(t) \leq n\}$ y, por consiguiente,

$$\mathbf{P}_\theta\left(\sum_{i=1}^{n} x_i \geq 1\right) = \sum_{k=0}^{n-1} e^{-\theta t}\frac{(\theta t)^k}{k!}. \tag{28}$$

Por eso cuando $t = n + 2\sqrt{n}$,

$$\alpha_1(\delta_c) = \mathbf{P}_1\left(\sum_{i=1}^{n} x_i \geq t\right) = \sum_{k=0}^{n-1} e^{-t}\frac{t^k}{k!},$$

$$\alpha_2(\delta_c) = \mathbf{P}_{\theta_2}\left(\sum_{i=1}^{n} x_i < t\right) = 1 - \sum_{k=0}^{n-1} e^{-\theta_2 t}\frac{(\theta_2 t)^k}{k!}.$$

Precisamente estas igualdades fueron utilizadas para calcular los valores exactos de $\alpha_i(\delta_c)$, $i = 1, 2$.

Nótese que a la par con (28) también se pueden escribir otras fórmulas para la distribución de $\sum_{i=1}^{n} x_i$, basadas en el hecho de que $\sum_{i=1}^{n} x_i \in \Gamma_{\theta,n}$.

5. Relación entre el c.m.p. y la eficacia asintótica de la e.v.m. Utilizando los cálculos realizados y los resultados de los §§ 1 y 2, ahora podemos demostrar el teorema 2.25.3 de la eficacia asintótica de la e.v.m. θ^* en la clase $\bar{K}^\circ$ de estimaciones asintóticamente centrales (la pertenencia de $\hat{\theta}^* \in \bar{K}^\circ$ ha sido establecida en el apartado (2.29.3).

Demostración del teorema 2.25.3. Admitamos lo contrario, es decir, el hecho de que existe una estimación asintóticamente normal θ^* tal que, para cualquier θ_1,

$$\lim_{n\to\infty} \mathbf{M}_{\theta_1} n(\theta^* - \theta_1)^2 = \sigma^2(\theta_1) < I^{-1}(\theta_1) = \lim_{n\to\infty} \mathbf{M}_{\theta_1} n(\hat{\theta}^* - \theta_1)^2.$$

Examinemos el problema de verificación de la hipótesis $H_1 = \{X \in \mathbf{P}_{\theta_1}\}$ frente a $H_2 = \{X \in \mathbf{P}_\theta,\ \theta = \theta_1 + vn^{-1/2}\}$ y construyamos para esto el criterio δ que tiene la forma siguiente:

$$\delta(X) = \begin{cases} H_1 & \text{si } \theta^* \leq \theta_1 + vn^{-1/2}, \\ H_2 & \text{si } \theta^* > \theta_1 + vn^{-1/2}, \end{cases}$$

donde hemos tomado, para precisar, que $v > 0$. Entonces

$$\alpha_1(\delta) = \mathbf{P}_{\theta_1}(\theta^* > \theta_1 + vn^{-1/2}) = \mathbf{P}_{\theta_1}\left(\frac{(\theta^* - \theta_1)\sqrt{n}}{\sigma(\theta_1)} > \frac{v}{\sigma(\theta_1)}\right) \to 1 - \Phi\left(\frac{v}{\sigma(\theta_1)}\right).$$

A continuación, la pertenencia de $\theta^* \in \bar{K}^\circ$ significa que

$$\alpha_2(\delta) = \mathbf{P}_\theta(\theta^* \leq \theta_1 + vn^{-1/2}) = \mathbf{P}_\theta(\theta^* \leq \theta) \to 1/2.$$

Examinemos ahora otro criterio $\delta_0(X)$ con la región crítica

$$\hat{\theta}^* - \theta_1 > (v + \gamma)/\sqrt{n}, \quad \gamma > 0,$$

que, como hemos establecido, será el c.a.m.p. (véase (21)). En vista de que cuando es bastante pequeña $\gamma > 0$,

$$(\upsilon + \gamma)\sqrt{I(\theta_1)} < \upsilon/\sigma(\theta_1),$$

para este criterio,

$$\lim_{n\to\infty} \alpha_1(\delta_0) = 1 - \Phi((\upsilon + \gamma)\sqrt{I(\theta_1)}) > 1 - \Phi(\upsilon/\sigma(\theta_1)),$$

$$\lim_{n\to\infty} \alpha_2(\delta_0) = \lim_{n\to\infty} \mathbf{P}_\theta(\theta^* \leqslant \theta + \gamma/\sqrt{n}) > 1/2.$$

Esto significa que a partir de cierto n, el criterio δ será mejor que el c.m.p. La contradicción obtenida demuestra el teorema. ◁

§4. Verificación de las hipótesis compuestas. Clases de criterios óptimos.

1. Planteamiento del problema y conceptos principales. En los §§ 1 y 2 hemos examinado los problemas menos complejos de verificación de las hipótesis cuando estas últimas son simples. Sin embrago, a menudo las hipótesis sujetas a verificación tienen una naturaleza más compleja. En el caso paramétrico, por ejemplo, la hipótesis puede tener la forma $\{X \in \mathbf{P}_\theta;\ \theta \in \Theta_1\}$, donde Θ_1 es un subconjunto dado del conjunto Θ. Evidentemente, tal hipótesis ya no define de manera unívoca la distribución de la muestra.

Llamaremos *compuesta* toda hipótesis H que no sea simple.

Por ejemplo, las hipótesis $\{X \in \Phi_{0,\sigma^2};\ \sigma \geqslant 0\}$, $\{X \in \Phi_{\alpha,1};\ \alpha \geqslant 0\}$ son compuestas.

Posteriormente en este capítulo examinaremos siempre los problemas relacionados con la verificación de *dos* hipótesis que designaremos por H_1 y H_2. Además, en los párrafos inmediatos nos limitamos a estudiar el *caso paramétrico* $X \in \mathbf{P}_\theta$, $\theta \in \Theta$. En este caso, las hipótesis H_i se pueden escribir de la forma siguiente:

$$H_i = \{X \in \mathbf{P}_\theta;\ \theta \in \Theta_i\},\quad \Theta_i \subset \Theta,\quad \Theta_1 \cap \Theta_2 = \varnothing.$$

Como los demás valores de θ que no pertenecen a $\Theta_1 \cup \Theta_2$ no se examinan en general, entonces, sin limitar la generalidad, podemos considerar que $\Theta = \Theta_1 \cup \Theta_2$, y que H_2 es una hipótesis *adicional* (o contraria) a H_1, así que la hipótesis H_2 también puede ser escrita en forma de $H_2 = \{H_1$ no es cierta$\}$. Al igual que en el § 2, una de las hipótesis será llamada *fundamental* (en este caso es H_1), y las hipótesis simples $H_\theta = \{X \in \mathbf{P}_\theta\}$, $\theta \in \Theta_2$ se llamarán *alternativas*.

La separación de una hipótesis fundamental entre todas las demás, a menudo refleja la actitud del investigador hacia el objeto de estudio. La hipótesis fundamental suele corresponder a cierta concepción, y la alternativa, a las desviaciones de ésta, cuya presencia ha de ser demostrada o

rechazada. Por regla general sólo hay una o un pequeño número de hipótesis fundamentales y una gran cantidad de hipótesis alternativas.

El procedimiento de admisión de las hipótesis se basa en el criterio estadístico. Como sólo examinamos dos hipótesis, entonces, al igual que en el § 2, todo criterio (randomizado) π será unívocamente definido por la función medible $\pi(x)$, $0 \leqslant \pi(x) \leqslant 1$, la cual determina la probabilidad de aprobación $\pi(X)$ de la hipótesis H_2 para cada muestra X (la realización de la elección aleatoria con probabilidad $\pi(X)$ debe llevarse a cabo con ayuda de un dispositivo adicional). Al igual que en el § 2, la función $\pi(x)$ se llama *crítica*. Para el criterio no randomizado δ, la función $\pi(x) = \delta(x)$ sólo adopta dos valores: 0 y 1; la región Ω_2 del espacio $\mathscr{X}^n$, en la que $\delta(x) = 1$ (región de admisión de H_2), en este caso se denomina región *crítica* y a menudo se identifica con el criterio δ.

Definición 1. Se llama *dimensión* o *probabilidad del error de primer género* del criterio π el número

$$\alpha_1(\pi) = \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi(X).$$

Es evidente que para los criterios no randomizados,

$$\alpha_1(\delta) = \sup_{\theta \in \Theta_1} \mathbf{P}_\theta(X \in \Omega_2).$$

Esta es la máxima probabilidad (respecto a $\theta \in \Theta_1$) de rechazar la hipótesis H_1 cuando ella es verdaderamente cierta. Por lo general, para facilitar las búsquedas de los criterios óptimos se examinan los criterios π que satisfacen la condición

$$\alpha_1(\pi) = \varepsilon \quad (\text{o } \alpha_1(\pi) \leqslant \varepsilon).$$

Designemos por K_ε la clase de tales criterios.

Llamaremos *nivel (de significación)**) del criterio π el número $1 - \alpha_1(\pi) = 1 - \varepsilon$.

La utilización del criterio $\delta \in K_\varepsilon$, estadísticamente significa que en una larga serie de experimentos para verificar la hipótesis H_1 con ayuda del criterio $\delta \in K_\varepsilon$, no nos equivocaremos más a menudo que en una porción de casos ε, si realmente era cierta la hipótesis H_1.

---

*) Con frecuencia se llama nivel de significación el número ε, y no el $1 - \varepsilon$. Pero esto es algo perverso: pues es natural considerar que cuando más alto sea el nivel de significación, tanto más "significativo" será el criterio. Partiendo precisamente de estas consideraciones hemos definido el nivel de significación (o de confianza) para los intervalos confidenciales. Como entre los criterios estadísticos y los intervalos confidenciales existe una relación directa (véase el § 8), no sería razonable cambiar esta terminología (al pasar a los criterios) por una contraria.

La elección del nivel de significación del criterio es, en gran medida, arbitraria. En calidad de ε se elige, de ordinario, uno de los valores estándar, tales como 0,005, 0,01, 0,05, 0,1. Esta estandarización tiene la ventaja de que permite reducir el volumen de las tablas que el estadista utiliza en su trabajo. No hay ninguna otra causa especial para escoger precisamente estos valores. Eligiendo el nivel de significación del criterio π, es necesario prestar atención a la *potencia* del criterio

$$\beta_\pi = \mathbf{M}_\theta \pi(X), \quad \theta \in \Theta_2.$$

Si ésta resulta demasiado pequeña, conviene, tal vez, sustituir el nivel $1 - \varepsilon$ por uno menor.

Nuestra actitud hacia la hipótesis antes de realizar el experimento es una circunstancia importante que puede influir en la elección del nivel de significación. Si creemos firmemente en la veracidad de la hipótesis (la probabilidad a priori $Q(H_1)$ en el planteamiento bayesiano del problema es grande), se necesitarán pruebas convincentes contra ella para que renunciemos a nuestra seguridad. En estas condiciones hacen falta criterios de alto nivel, y ε se elige muy pequeño (entonces, la toma de un valor perteneciente a Ω_2 será demasiado inverosímil si es cierta H_1).

Aquí se utiliza la misma concepción que hemos expuesto al construir los intervalos confidenciales. La misma consiste en lo siguiente: si la probabilidad ε de cierto suceso A es pequeña, consideraremos prácticamente imposible el hecho de que este suceso ocurra al realizar una sola prueba.

Entre algunos especialistas de estadística matemática también existe otro punto de vista, el cual radica en que no hay necesidad de asignar un nivel de significación fijo y que para su elección preliminar no hay una regla razonable. Ellos consideran la verificación de las hipótesis no como un procedimiento que conduce obligatoriamente a la aprobación de una de dos hipótesis, sino como cierto proceso que se desarrolla en la conciencia del investigador y que determina la actitud de éste hacia las hipótesis. Desde este punto de vista, al número de significación registrado se le puede anteponer el nivel "realmente alcanzable" que se determina del modo siguiente. Examinemos la familia de criterios no randomizados δ de nivel $1 - \varepsilon$ cuando ε recorre los valores del intervalo (0, 1), y designemos por $\Omega_{2,\varepsilon}$ la región crítica δ, suponiendo que $\Omega_{2,\varepsilon_2} \subset \Omega_{2,\varepsilon_1}$ cuando $\varepsilon_2 < \varepsilon_1$.

Definición 2. Llámase *nivel realmente alcanzable* de la familia de criterios δ en la muestra X, la variable aleatoria $1 - \varepsilon(X)$, donde

$$\varepsilon(X) = \inf \{\varepsilon \colon X \in \Omega_{2,\varepsilon}\}.$$

Cuanto mayor es $1 - \varepsilon(X)$ tanto más fuertemente testimonia la muestra contra la hipótesis H_1.

El valor $\varepsilon(X)$ da la posibilidad de aceptar o rechazar la hipótesis para cualquiera que sea el nivel $1 - \varepsilon$ dado de antemano, mediante la simple comparación de $\varepsilon(X)$ con ε.

Ejemplo 1. En el párrafo anterior hemos construido el c.m.p. para verificar la hipótesis $H_1 = \{X \in \Gamma_{1,1}\}$ frente a la hipótesis $H_2 = \{X \in \Gamma_{1/2,1}\}$. Este criterio tiene la siguiente región crítica:

$$\Omega_2 = \left\{x \in \mathscr{X}^n:\ \sum_{i=1}^{n} (x_i - 1) > d_1\right\}.$$

Supongamos que para la muestra X de volumen $n = 10$ ha resultado $\sum_{i=1}^{10} x_i = 18$. Como para la hipótesis H_1 $\sum_{i=1}^{n} x_i \in \Gamma_{1,n}$ y $\Gamma_{1,n}((a, b)) = \mathbf{H}_{2n}((2a, 2b))$, entonces $\Gamma_{1,10}((18, \infty)) = \mathbf{H}_2((36, \infty)) = 0{,}0154$ (véanse las tablas III ó [8], y el nivel que en este caso se alcanza realmente será igual a $1 - \varepsilon(X) = 1 - 0{,}0154 = 0{,}9846$, así que la hipótesis H_1 será rechazada por el c.m.p. de nivel $1 - \varepsilon = 0{,}98$ y no será rechazada por el c.m.p. de nivel $1 - \varepsilon = 0{,}99$.

2. Criterios uniformemente más potentes. Volvamos a examinar los criterios randomizados arbitrarios π que hemos acordado designarlos por la función crítica $\pi(x)$, $x \in \mathscr{X}^n$ (La función $\pi(x)$ también se puede llamar función estadística (randomizada) de decisión).

Si existe una estadística suficiente $S(X)$, entonces es posible limitarse a los criterios $\pi(X)$ que dependen de X sólo por la estadística suficiente $S(X)$, o sea, por los criterios representables en la forma $\pi(X) = \varphi(S(X))$. Pues sabemos que toda la información sobre el parámetro desconocido está concentrada en S, y la utilización de otras estadísticas (otra información sobre la muestra X) no tiene sentido.

Como ya hemos señalado, para determinar los criterios óptimos, se reduce, de ordinario, el conjunto de criterios que se examinan, hasta la clase K_ε de los criterios de nivel registrado. Entre ellos se puede tratar de hallar un criterio tal, para el que la potencia

$$\beta_\pi(\theta) = \mathbf{M}_\theta \pi(X)$$

en la región Θ_2 sea máxima (es decir, la probabilidad del error de segundo género $1 - \beta_\pi(\theta)$ debe ser mínima). Con otras palabras, ha de ser máxima la probabilidad de aceptar la hipótesis H_2 cuando ésta es cierta.

La función $\beta_\pi(\theta) = \mathbf{M}_\theta \pi(X)$ también suele llamarse *función de potencia* del criterio π.

Definición 3. El criterio $\pi^\circ \in K_\varepsilon$ se denomina *criterio uniformemente más potente* (c.u.m.p.) en K_ε, si para cualquier $\pi \in K_\varepsilon$

$$\beta_{\pi^\circ}(\theta) \geqslant \beta_\pi(\theta) \quad \text{para todos} \quad \theta \in \Theta_2. \tag{1}$$

Claro está que c.u.m.p. existe no siempre, ni mucho menos. Si tal criterio π° existiera, la función de potencia $\beta_{\pi^\circ}(\theta)$ para él en el gráfico permanecería más alta que cualquier otra función $\beta_\pi(\theta)$ en la región Θ_2 a condición

de que ambas funciones no excedan el valor ε en la región Θ_1 (pues $\alpha_1(\pi) = \sup_{\theta\in\Theta_1} \beta_\pi(\theta)$), así que $\beta_{\pi^\circ}(\theta)$ es la envolvente de la familia $\{\beta_\pi(\theta)\}$ en la región Θ_2.

Supongamos que $\Theta_1 = \{\theta_1\}$, $\mathbf{M}_{\theta_1}\pi^\circ(X) = \varepsilon$. Entonces el c.u.m.p. π° será, evidentemente, el c.m.p. de nivel $1 - \varepsilon$ para verificar la hipótesis $\{\theta = \theta_1\}$ frente a la alternativa $\{\theta = \theta_2\}$ con cualquier $\theta_2 \in \Theta_2$. Como conocemos la forma del c.m.p., de aquí surge el siguiente procedimiento natural de búsqueda del c.u.m.p.: lo encontraremos si resulta que en el problema antes planteado, acerca de la verificación de las hipótesis $\{\theta = \theta_1\}$ y $\{\theta = \theta_2\}$, el c.m.p. no depende de θ_2.

También es cierto lo contrario: si el c.m.p. de K_ε para verificar la hipótesis $\{\theta = \theta_1\}$ frente a $\{\theta = \theta_2\}$, $\theta_2 \in \Theta_2$ depende considerablemente de θ_2, esto significará que el c.u.m.p. para verificar $\{\theta = \theta_1\}$ frente a $\theta \in \Theta_2$ no existe.

Si la hipótesis H_2 es simple (Θ_2 consta de un solo punto θ_2), el concepto de c.u.m.p. pierde parcialmente su sentido y se transforma en concepto de c.m.p. ordinario, o sea, en un criterio para el que en la clase K_ε se maximiza $\mathbf{M}_{\theta_2}\pi(X)$.

Definamos ahora los criterios bayesianos y minimax para comprobar las hipótesis compuestas.

3. Criterios bayesianos. Al comprobar las hipótesis compuestas distinguiremos dos enfoques bayesianos.

a) *Enfoque bayesiano completo.* Consiste en la suposición de que las hipótesis $H_\theta = \{X \in \mathbf{P}_\theta\}$, $\theta \in \Theta$ se escogen al azar, con una distribución a priori $\mathbf{Q}$. Con otras palabras, en $\Theta = \Theta_1 \cup \Theta_2$ se registra cierta σ-álgebra de los subconjuntos $\mathfrak{S}$, $\Theta_1 \in \mathfrak{S}$, $\Theta_2 \in \mathfrak{S}$, y θ se considera como variable aleatoria en el espacio muestral $(\Theta, \mathfrak{S}, \mathbf{Q})$.

La distribución $\mathbf{Q}$ induce la distribución $\mathbf{Q}_i$ en Θ_i, $i = 1, 2$ y las probabilidades $q_i = \mathbf{Q}(\theta \in \Theta_i)$, así que $\mathbf{Q} = q_1 Q_1 + q_2 \mathbf{Q}_2$. La hipótesis de que $\theta \in \Theta_i$ se elige al azar, con una distribución $\mathbf{Q}_i$, la designaremos por H_{Q_i}.

Definición 4. El criterio π_Q se llama *bayesiano* si es un criterio bayesiano correspondiente a la distribución a priori (q_1, q_2) para verificar dos hipótesis simples H_{Q_1} y H_{Q_2} (véase el § 1).

b) *Enfoque parcialmente bayesiano.* Aquí se supone que han sido dadas las distribuciones a priori $\mathbf{Q}_i$ en Θ_i, pero que faltan las probabilidades a priori q_1, q_2. En este caso se trata de la verificación de dos hipótesis simples H_{Q_1} y H_{Q_2}.

Designemos, como antes,

$$K_\varepsilon = \{\pi: \sup_{\theta\in\Theta_1} \mathbf{M}_\theta \pi(X) \leqslant \varepsilon\}$$

y pongamos

$$K_\varepsilon^{Q_1} = \{\pi: \mathbf{M}_{Q_1}\pi(X) \leqslant \varepsilon\},$$

donde $\mathbf{M}_{Q_i}$ designa la esperanza matemática incondicional de la distribución en $\Theta_i \times \mathscr{X}^n$, engendrada por $\mathbf{Q}_i$ y $\mathbf{P}_\theta$.

Definición 5. El criterio π_{Q_1,Q_2} se llama *bayesiano* en $K_\varepsilon^{Q_1}$ si es el c.m.p. de nivel $1-\varepsilon$ para la verificación de dos hipótesis simples H_{Q_1} y H_{Q_2}.

Si una de las hipótesis H_i degenera en hipótesis simple (Θ_1 ó Θ_2 unipuntualmente), también degenerá la distribución respectiva. En este caso acortaremos el índice en la designación π_{Q_1,Q_2} y escribiremos π_{Q_1} en vez de π_{Q_1,Q_2} si $\Theta_2 = \{\theta_2\}$ unipuntualmente.

La construcción de los criterios π_{Q_1,Q_2} no presenta dificultades. Utilizaremos estos criterios como medio auxiliar para construir los c.u.m.p. y los minimax.

4. Criterios minimax.

Definición 6. El criterio $\bar{\pi}$ para verificar $H_1 = \{\theta \in \Theta_1\}$ frente a $H_2 = \{\theta \in \Theta_2\}$ se llama *minimax* en K_ε (en $K_\varepsilon^{Q_1}$) si $\bar{\pi} \in K_\varepsilon (\bar{\pi} \in K_\varepsilon^{Q_1})$, y para él se maximiza

$$\inf_{\theta \in \Theta_2} \mathbf{M}_\theta \pi(X) = \inf_{\theta \in \Theta_2} \beta_\pi(\theta).$$

Sería más correcto llamar este criterio *maximín* (se maximiza el mínimo). Sin embargo, a pesar de todo utilizaremos el término único "minimax", ya que el mismo conserva su sentido aún cuando se trata no de la potencia, sino de las probabilidades de segundo género.

Los criterios bayesianos y minimax se examinan más detalladamente en el § 9. Los párrafos están dedicados a la aclaración de las condiciones en las que es posible construir los c.u.m.p.

§ 5. Criterios uniformemente más potentes

En este párrafo examinaremos dos importantes casos particulares, referentes al parámetro unidimensional θ cuando se logra construir el c.u.m.p. También obtendremos un resultado útil en cuanto a la construcción del c.m.p.

1. Alternativas unilaterales. Relación monótona de verosimilitud. Supongamos que la hipótesis fundamental H_1 consiste en que $\theta \leqslant \theta_1$, y la hipótesis alternativa H_2, en que $\theta > \theta_1$. Llamaremos *unilateral* tal hipótesis H_2, a distinción, digamos, de la hipótesis $H_2 = \{\theta \neq \theta_1\}$ (adicional a $H_1 = \{\theta = \theta_1\}$), la cual es bilateral, puesto que admite desviaciones respecto a θ_1 en ambas direcciones.

Nuestra otra suposición consiste en lo siguiente. Supongamos que se cumple la condición (A_0) y que existe una función $T(x)$ tal, que para todos θ, θ_0, $\theta > \theta_0$, la relación de verosimilitud

$$\frac{f_\theta(x)}{f_{\theta_0}(x)} \tag{1}$$

es una función no decreciente (o no creciente) de $T(x)$. En este caso se dice que la familia $\{\mathbf{P}_\theta\}$ tiene una *relación de verosimilitud monótona.*

En vista de que T es una estadística suficiente, entonces $f(x) = \psi(T(x), \theta)h(x)$, y la condición enunciada corresponderá a la relación $\psi(T, \theta)/\psi(T, \theta_0)$. Esta condición significa que para todos $\theta > \theta_0$ y para cualquier $d > 0$, la desigualdad $f_\theta(x)/f_{\theta_0}(x) \geqslant d$ será resoluble en la forma $T(x) \geqslant c_n(\theta, \theta_0, d)$ (o bien $T(x) \geqslant c_n(\theta, \theta_0, d)$).

Por ejemplo, las familias $\{\Phi_{\alpha,1}\}$ y $\{\Phi_{0,\sigma^2}\}$ tienen una relación de verosimilitud monótona, ya que

$$\frac{f_{\alpha,1}(X)}{f_{\alpha_0,1}(x)} = \exp\left\{(\alpha - \alpha_0)n\bar{x} - \frac{n}{2}(\alpha^2 - \alpha_0^2)\right\},$$

$$\frac{f_{0,\sigma^2}(X)}{f_{0,\sigma_0^2}(X)} = \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\sigma_0^2}\right)\sum_{i=1}^{n} x_i^2\right\},$$

y las desigualdades respectivas tendrán la forma ($\alpha > \alpha_0$, $\sigma > \sigma_0$)

$$\bar{x} \geqslant c_n(\alpha, \alpha_0, d) = \frac{1}{2}(\alpha + \alpha_0) + \frac{\ln d}{n(\alpha - \alpha_0)} \quad (T(X) = \bar{x}),$$

$$\sum_{i=1}^{n} x_i^2 \leqslant c_n(\sigma, \sigma_0, d) = \frac{2(\sigma\sigma_0)^2}{\sigma^2 - \sigma_0^2} \ln d \left(T(X) = \sum_{i=1}^{n} x_i^2\right).$$

Muchas familias paramétricas del § 2.2 también tienen una relación de verosimilitud monótona. En lo sucesivo, para precisar, consideramos que (61) es una función *no decreciente* $T(x)$.

Teorema 1. *Sea θ un parámetro unidimensional y supongamos que $\{\mathbf{P}_\theta\}$ tiene una relación de verosimilitud monótona. Entonces*

1) *En K_ε existe c.u.m.p. para verificar la hipótesis $H_1 = \{\theta \leqslant \theta_1\}$ frente a la alternativa $H_2 = \{\theta > \theta_1\}$, el cual tiene la forma siguiente:*

$$\pi^\circ(X) = \begin{cases} 1, & si \quad T(X) > c, \\ p, & si \quad T(X) = c, \\ 0, & si \quad T(X) < c, \end{cases} \tag{2}$$

donde c y p se deducen de la condición

$$\mathbf{M}_{\theta_1}\pi^\circ(X) = \mathbf{P}_{\theta_1}(T(X) > c) + p\mathbf{P}_{\theta_1}(T(X) = c) = \varepsilon. \tag{3}$$

2) *La función de potencia $\beta^\circ(\theta) = \mathbf{M}_\theta\pi^\circ(X)$ crece estrictamente en θ con todos θ para los cuales $\beta^\circ(\theta) < 1$.*

3) *Con todos θ_0 el criterio (2) es el c.u.m.p. en la clase $K_{\beta^\circ(\theta_0)}$ para verificar la hipótesis $H_1^0 = \{\theta \leqslant \theta_0\}$ frente a $H_2^0 = \{\theta > \theta_0\}$.*

4) *Para cualquier $\theta < \theta_1$, nuestro criterio minimiza $\beta(\theta) = \mathbf{M}_\theta\pi(X)$ en la clase K_ε.*

Demostración. Examinemos primeramente las hipótesis simples $\{\theta = \theta_1\}$ y $\{\theta = \theta_2\}$, $\theta_2 > \theta_1$. El c.m.p. para verificar estas hipótesis en la clase de criterios π, para los cuales $\mathbf{M}_{\theta_1}\pi(X) = \varepsilon$, tiene, según el teorema 2.1, la forma (2), ya que la desigualdad $Z(X) > d$ equivale a $T(X) > c$ (en caso de la debida correspondencia entre c y d), donde las constantes c y p se deducen de (3) (compárese con (2.3)). Como los números c y p de la ecuación de forma (3) se determinan de un modo único, entonces también obtenemos que el criterio (2) será el c.m.p. en $K_{\beta^\circ(\theta_0)}$ para verificar la hipótesis $\{\theta = \theta_0\}$ frente a $\{\theta = \theta_2\}$, $\theta_2 > \theta_0$. De aquí y del teorema 2.1 (véase (2.4) resulta que $\beta^\circ(\theta_2) > \beta^\circ(\theta_0)$.

Como $\beta^\circ(\theta)$ no decrece, entonces

$$\mathbf{M}_\theta \pi^\circ(X) \leqslant \varepsilon \quad \text{cuando} \quad \theta \leqslant \theta_1. \tag{4}$$

La clase K_ε de los criterios π que satisfacen (4) está presente en la clase $\{\pi: \mathbf{M}_{\theta_1}\pi(X) = \varepsilon\}$. En vista de que el criterio (2) maximiza $\beta(\theta_2)$ en esta última clase, también maximizará $\beta(\theta_2)$ en K_ε. Queda señalar que el criterio (2) no depende de ningún modo de θ_2 y, por consiguiente, las conclusiones sacadas son válidas para cualquier $\theta_2 > \theta_1$. Aquí pues, han sido demostradas las primeras tres afirmaciones del teorema.

La cuarta afirmación se deduce de las tres primeras si éstas se aplican al problema de la verificación de la hipótesis $H_1' = \{\theta \geqslant \theta_1\}$ frente a $H_2' = \{\theta < \theta_1\}$, para la cual el c.u.m.p. en la clase $\{\Pi(X): \mathbf{M}_\theta\Pi(X) \leqslant 1 - \varepsilon, \theta \geqslant \theta_1\}$ tendrá la forma $\Pi^\circ(X) = 1 - \pi^\circ(X)$, y la función $1 - \beta^\circ(\theta) = \mathbf{M}_\theta\Pi^\circ(X)$ en máxima función de potencia cuando $\theta < \theta_1$. ◁

Una importante clase de familias de distribuciones que admiten la relación de verosimilitud monótona es formada por la familia exponencial monoparamétrica (véase el § 2.15) cuando la densidad $f_\theta(x)$ es representable en la forma

$$f_\theta(x) = h(x)\exp\{a(\theta)U(x) + V(\theta)\}. \tag{5}$$

En efecto, en este caso

$$\frac{f_\theta(x)}{f_{\theta_0}(x)} = \exp\left\{(a(\theta) - a(\theta_0))\sum_{i=1}^{n} U(x_i) + n(V(\theta) - V(\theta_0))\right\}$$

y la relación de verosimilitud dependerá monótonamente de $T(x) = \sum_{i=1}^{n} U(x_i)$ si $a(\theta) - a(\theta_0)$ conserva el signo en todos θ, θ_0, $\theta > \theta_0$.

Corolario 1. *Supongamos que $f_\theta(x)$ tiene la forma* (5), *donde $a(\theta)$ es una función monótona. Entonces existe el c.u.m.p. π° en la clase K_ε para la verificación de la hipótesis $H_1 = \{\theta \leqslant \theta_1\}$ frente a $H_2 = \{\theta > \theta_1\}$. Si $a(\theta)$ crece, este criterio tiene la forma* (2) y (3). *Si $a(\theta)$ decrece, las desigualdades en* (2) y (3) *se sustituyen por las contrarias.*

Nótese que si se verifica la alternativa bilateral, por ejemplo, la hipótesis $H_1 = \{\theta = \theta_1\}$ frente a $H_2 = \{\theta \neq \theta_1\}$, entonces el c.u.m.p. para la familia exponencial (5) ya no existe. En efecto, admitamos, para abreviar, que $a(\theta)$ crece y que $\mathbf{P}_\theta$-distribución de $T(X) = \sum_{i=1}^{n} U(x_i)$ para todos θ es absolutamente continua. Entonces, de acuerdo con el teorema 2.1, el c.m.p. para la verificación de $\{\theta = \theta_1\}$ frente a $\{\theta = \theta_2\}$ será no randomizado y tendrá la región crítica $T(X) \geqslant c$ si $\theta_2 > \theta_1$. No obstante, si $\theta_2 < \theta_1$, la región crítica tendrá la forma $T(X) < c$. Vemos que la potencia máxima en el punto θ_2 se alcanzará con criterios muy diferentes en función del signo de diferencia de $\theta_2 - \theta_1$. Del teorema 1 se deduce que si tomamos cualquiera de estos criterios, por ejemplo, aquél para el cual $\pi(X) = 1$ cuando $T(X) \geqslant c$, entonces éste será el c.u.m.p. para todos $\theta_2 > \theta_1$ y a ciencia cierta no será tal para $\theta_2 < \theta_1$.

Ya hemos señalado que la situación de dos hipótesis simples en el teorema 2.1. del c.m.p. es, en cierto sentido, simétrica (el c.m.p. minimiza la probabilidad del error de segundo género $\alpha_2(\pi)$ si ha sido registrado el valor de $\alpha_1(\pi)$ y, al contrario, minimiza $\alpha_1(\pi)$ si se ha registrado $\alpha_2(\pi)$. En el planteamiento del problema de la verificación de las hipótesis compuestas no existe tal simetría. Con esta circunstancia está vinculado el siguiente hecho interesante. Acabamos de ver que para una familia exponencial no existe el c.u.m.p. destinado a verificar la hipótesis $H_1 = \{\theta = \theta_1\}$ frente a $H_2 = \{\theta \neq \theta_1\}$. De las investigaciones realizadas es fácil comprender que no existe tampoco el c.u.m.p. para la verificación de la hipótesis $\{\theta_1 < \theta < \theta_2\}$ frente a la alternativa $\{\theta \notin (\theta_1 \theta_2)\}$. No obstante, si examinamos ahora, en calidad de hipótesis fundamental H_1, la $H_1 = \{\theta \notin (\theta_1, \theta_2)\}$, y en calidad de alternativa, la hipótesis $H_2 = \{\theta \in (\theta_1, \theta_2)\}$, entonces el c.u.m.p. en K_ε ya existirá. Así pues, vamos examinar ahora la segunda posibilidad cuando se logra construir el c.u.m.p.

2. Hipótesis fundamental bilateral. Familia exponencial.

Teorema 2. *Supongamos que $f_\theta(x)$ se define por la igualdad* (5) *y que se verifica la hipótesis $H_1 = \{\theta \notin (\theta_1, \theta_2)\}$, $\theta_1 < \theta_2$,* frente a la alternativa *$H_2 = \{\theta \notin (\theta_1, \theta_2)\}$. En este caso si la función $a(\theta)$ es monótona,*

1) *en la clase $K_\varepsilon = \{\pi: \sup_{\theta \notin (\theta_1, \theta_2)} \mathbf{M}_\theta \pi(X) \leqslant \varepsilon\}$ existe un c.u.m.p. π° que tiene la forma*

$$\pi^\circ(x) = \begin{cases} 1, & si \quad c_1 < T(x) < c_2, \\ p_i, & si \quad T(x) = c_i, \quad i = 1, 2, \\ 0, & si \quad T(x) \notin [c_1, c_2], \end{cases} \tag{6}$$

donde $T(x) = \sum_{i=1}^{n} U(x_i)$ y las constantes c_i, p_i se deducen de las condiciones

$$\mathbf{M}_{\theta_1} \pi^\circ(X) = \mathbf{M}_{\theta_2} \pi^\circ(X) = \varepsilon. \tag{7}$$

2) *Este criterio maximiza la función de potencia* $\beta(\theta) = \mathbf{M}_\theta \pi(X)$ *a condición de* (7) *dentro del intervalo* (θ_1, θ_2), *y la minimiza fuera de este intervalo* (véase la fig. 4).

3) *Cuando* $0 < \varepsilon < 1$, *la función* $\beta^\circ(\theta)$ *tiene el máximo en cierto punto* $\theta_0 \in (\theta_1, \theta_2)$ *y decrece estrictamente al alejarse* θ *de* θ_0 *a la derecha o a la izquierda. Además excluimos el caso cuando la distribución de* $T(X)$ *está concentrada en dos puntos, o sea, cuando existen tales* t_1, t_2 que

$$\mathbf{P}_\theta(T(X) = t_1) + \mathbf{P}_\theta(T(X) = t_2) = 1 \quad \textit{para todos } \theta. \tag{8}$$

En las investigaciones que se realizan también es útil la afirmación siguiente.

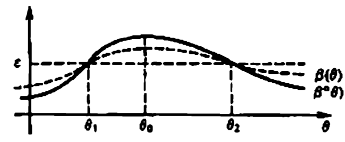


Fig. 4. Forma de la función de potencia $\beta^0(\theta) = \mathbf{M}_\theta \pi^0(X)$ y $\beta(\theta) = \mathbf{M}_\theta \pi(X)$ para el criterio arbitrario $\pi \in K$.

Lema 1. *Las ecuaciones* (7) *para* $0 < \varepsilon < 1$ *son siempre resolubles con respecto a* c_i y p_i, $i = 1, 2$.

La demostración de este lema se dará más tarde.

Demostración del teorema 2. Escribamos la función de verosimilitud en la forma

$$f_\theta(x) = c(\theta) e^{a(\theta) T(x)} h(x), \tag{9}$$

donde, supondremos, con el fin de precisar, que $a(\theta)$ crece estrictamente.

Examinemos el siguiente planteamiento bayesiano del problema. Admitamos que se verifica la hipótesis fundamental "mixta" H, la cual consiste en que $\{\theta = \theta_1\}$ con probabilidad q, y $\{\theta = \theta_2\}$ con probabilidad $1 - q$ frente a la alternativa $H_0 = \{\theta = \theta_0\}$, $\theta_0 \in (\theta_1, \theta_2)$. Supongamos después, que las probabilidades a priori de las hipótesis H y H_0 son iguales a r y $1 - r$, respectivamente. Como las hipótesis H y H_0 determinan por completo la distribución de la muestra, ellas pueden considerarse simples y podemos hacer uso de los resultados del § 2. En este caso el criterio bayesiano (designémoslo por π°) tendrá la forma

$$\pi^\circ(X) = \begin{cases} 1, & \text{si} \quad R(X) \equiv \dfrac{f_{\theta_0}(X)}{q f_{\theta_1}(X) + (1 - q) f_{\theta_2}(X)} > \dfrac{r}{1 - r}, \\ p(X), & \text{si} \quad R(X) = \dfrac{r}{1 - r} \\ 0, & \text{si} \quad R(X) < \dfrac{p}{1 - r}. \end{cases} \tag{10}$$

En virtud de (9) la desigualdad $R(X) > r/(1 - r)$ es equivalente a la desigualdad

$$q\,\frac{c(\theta_1)}{c(\theta_0)}\,e^{(a(\theta_1) - a(\theta_0))T} + (1 - q)\,\frac{c(\theta_2)}{c(\theta_0)}\,e^{(a(\theta_2) - a(\theta_0))T} < \frac{1 - r}{r}. \qquad (11)$$

Como $a(\theta_1) - a(\theta_0) < 0$, $a(\theta_2) - a(\theta_0) > 0$, aquí el primer miembro es una función convexa de T. Esto quiere decir que (11) se puede escribir en la forma

$$c_1 < T < c_2,$$

donde $c_i = c_i(q, r)$; los números $c_1 < c_2$ recorren, al variar q y r, todos los valores posibles. La función $p(X)$ en (10) se supone igual a p_1 si $T(X) = c_1$ y p_2 si $T(X) = c_2$.

Según el lema 1, habrá c_i, $i = 1, 2$ (o bien, que es lo mismo, q y r) y p_i tales que (7) sea cumplida. Mostremos ahora, que la función $\pi^\circ(X)$ definida en (10) o, que es lo mismo, en 6, poseerá todas las propiedades enunciadas en el teorema 2. Lo dicho significa que ahora consideramos π° simultáneamente como función de decisión para la verificación de H_1 frente a H_2. Como el criterio π° es bayesiano (para la verificación de H frente a H_0), entonces, para cualquier otro criterio π,

$$r[q\mathbf{M}_{\theta_1}\pi^\circ + (1 - q)\mathbf{M}_{\theta_2}\pi^\circ] + (1 - r)\mathbf{M}_{\theta_0}(1 - \pi^\circ) \leqslant$$
$$\leqslant r[q\mathbf{M}_{\theta_1}\pi + (1 - q)\mathbf{M}_{\theta_2}\pi] + (1 - r)\mathbf{M}_{\theta_0}(1 - \pi). \qquad (12)$$

Por consiguiente, si el criterio π, a la par con π°, satisface (7), entonces

$$\mathbf{M}_{\theta_0}\pi^\circ \geqslant \mathbf{M}_{\theta_0}\pi.$$

Esto significa que en cada punto interior $\theta_0 \in (\theta_1, \theta_2)$, el criterio π° maximiza la función de potencia $\beta(\theta) = \mathbf{M}_\theta\pi$ en la clase de criterios π que satisface (7). Pero las condiciones (7) destacan una clase de criterios que es más amplia que K_ε. Por lo tanto, π° también maximizará $\beta(\theta)$ en K_ε. En vista de que el criterio π° no depende de θ_0, el mismo será el c.u.m.p. en K_ε.

También cabe señalar que, en virtud del teorema 2.1,

$$\beta^\circ(\theta_0) = \mathbf{M}_{\theta_0}\pi^\circ \geqslant \varepsilon$$

y aquí la igualdad sólo es posible en el caso de que

$$qf_{\theta_1}(x) + (1 - q)f_{\theta_2}(x) = f_{\theta_0}(x) \qquad (13)$$

μ^n casi por doquier.

De un modo absolutamente análogo podemos convencernos, con ayuda de (12), de que π° minimizará $\mathbf{M}_{\theta}\pi$ para $\mathbf{M}_{\theta_1}\pi$, $\mathbf{M}_{\theta_2}\pi$ registradas (aquí utilizamos las mismas consideraciones que en la demostración de los teoremas del § 1).

Mostremos ahora, que π° minimiza $\beta(\theta)$ fuera de (θ_1, θ_2). Sea $\theta^\circ < \theta_1$. Sustituyamos en las investigaciones precedentes, los tres puntos $(\theta_1, \theta_0, \theta_2)$ por los tres puntos $(\theta^\circ, \theta_1, \theta_2)$ y notemos que para el nuevo problema, el criterio π° volverá a ser bayesiano (pues su forma no depende de la elección de los puntos θ_i, $i = 0, 1, 2$) en la clase de criterios π para los cuales $\mathbf{M}_{\theta_0}\pi = \beta^\circ(\theta^\circ)$, $\mathbf{M}_{\theta_2}\pi = \varepsilon$. Pero, según la observación hecha anteriormente, π° minimizará $\mathbf{M}_{\theta_0}\pi$ para $\mathbf{M}_{\theta_1}\pi$ y $\mathbf{M}_{\theta_2}\pi$ registradas. Las primeras dos afirmaciones del teorema quedan demostradas.

Demostremos la tercera afirmación. Nótese previamente que, utilizando la sustitución de las variables de integración, podemos escribir

$$\mathbf{P}_\theta(T \in A) = c(\theta) \int\limits_{\{x:\, T(x) \in A\}} e^{a(\theta)T(x)} h(x)\mu^n(dx) = c(\theta) \int\limits_{t \in A} e^{a(\theta)t} \nu(dt),$$

donde la medida ν se define por la relación

$$\nu(A) = \int\limits_{\{x:\, T(x) \in A\}} h(x)\mu^n(dx).$$

Esto quiere decir que la distribución T respecto a la medida ν tiene densidad (véase también el lema 2.15.1) $g_\theta(t) = c(\theta)e^{a(\theta)t}$ y, por consiguiente, también pertenece a la familia exponencial. Luego, en virtud de la monotonía de $a(\theta)$ se puede introducir un nuevo parámetro $\beta = a(\theta)$ sin modificar absolutamente el problema y sus condiciones. Por consiguiente, podemos considerar, sin limitar la generalidad, que $a(\theta) = \theta$. En este caso las funciones $c(\theta) = \left[\int e^{\theta t}\nu(dt)\right]^{-1}$ y $\beta^\circ(\theta) = \mathbf{M}_\theta \pi^\circ(X)$ serán, evidentemente, continuas. Admitamos ahora que la afirmación del teorema acerca del carácter del comportamiento de $\beta^\circ(\theta)$ no es cierta. Entonces habrá tres puntos $\theta' < \theta'' < \theta'''$ para los cuales

$$\beta^\circ(\theta') = \beta^\circ(\theta'') = \beta^\circ(\theta''') = \alpha \in (0, 1). \tag{14}$$

Hemos visto que π° maximiza $\beta(\theta'')$ para las condiciones $\beta(\theta') = \beta(\theta''') = \alpha$, con la particularidad de que si no se cumple la condición que tiene la forma (13), entonces $\beta^\circ(\theta'') > \alpha$. Pero en nuestro caso la igualdad (13) quiere decir que

$$q\frac{f_{\theta'}}{f_{\theta''}} + (1-q)\frac{f_{\theta'''}}{f_{\theta''}} = q\frac{c(\theta')}{c(\theta'')}e^{(\theta'-\theta'')T} + (1-q)\frac{c(\theta''')}{c(\theta'')}e^{(\theta'''-\theta'')T} = 1$$

ν-casi por doquier. En virtud de la convexidad del primer miembro respecto a T, esta igualdad es posible no más que para dos valores de T. Por lo tanto, si (8) se excluye, entonces $\beta^\circ(\theta'') > \beta^\circ(\theta') = \alpha$, y (14) es imposible. ◁

La demostración del lema 1 se llevará a efecto suponiendo simplemente que la distribución $T(X)$ es continua, o sea, que $\mathbf{P}_\theta(T = c) = 0$ para todos θ y c. Esto nos liberará de complicaciones poco importantes. En este caso, en virtud de las observaciones hechas al final de la demostración del teorema 2, podemos escribir

$$\mathbf{M}_\theta\pi^\circ(X) = \mathbf{P}_\theta(T \in (c_1,\ c_2)) = \int_{c_1}^{c_2} g_\theta(t)\nu(dt) = c(\theta)\int_{c_1}^{c_2} e^{\theta t}\nu(dt).$$

Esta será una función continua de θ, c_1, c_2.

Designemos por c_+ el valor de c para el cual $\mathbf{P}_{\theta_1}(T \leqslant c_+) = 1 - \varepsilon$. Entonces, en $(-\infty,\ c_+)$ estará definida una función $d(c)$ tal, que

$$\mathbf{P}_{\theta_1}(T \in (c,\ d(c))) = \int_c^{d(c)} g_{\theta_1}(t\nu(dt) = \varepsilon.$$

Naturalmente que $d(c)$ es una función continua creciente.

Demostraremos la afirmación requerida si nos convencemos de que la función

$$\psi(c) = \mathbf{P}_{\theta_2}(T \in (c,\ d(c))) = \int_c^{d(c)} g_{\theta_2}(t)\nu(dt)$$

crece continuamente, $\psi(-\infty) < \varepsilon$, $\psi(c_+) > \varepsilon$. En este caso existirá un valor de c_0 tal, que $\psi(c_0) = \varepsilon$ y, por lo tanto $\mathbf{P}_{\theta_i}(c_0,\ d(c_0)) = \varepsilon$, $i = 1,\ 2$.

La continuidad de $\psi(c)$ es evidente. Demostremos ahora la monotonía. Escribamos $\psi(c)$ en la forma

$$\psi(c) = \int_c^{d(c)} g_{\theta_1}(t)r(t)\nu(dt), \tag{15}$$

donde $r(t)$ es la densidad de la $\mathbf{P}_{\theta_2}$-distribución de T respecto a la $\mathbf{P}_{\theta_1}$-distribución:

$$r(t) = \frac{c(\theta_2)}{c(\theta_1)}\, e^{(\theta_2 - \theta_1)t}.$$

Supongamos, para precisar, que Δ es tal, que $c + \Delta < d(c)$. En este caso, como

$$\int_c^{c+\Delta} g_{\theta_1}(t)\nu(dt) = \int_c^{d(c+\Delta)} g_{\theta_1}(t)\nu(dt), \tag{16}$$

entonces

$$\psi(c + \Delta) - \psi(c) = \int_c^{d(c+\Delta)} g_{\theta_1}(t)r(t)\nu(dt) - \int_c^{c+\Delta} g_{\theta_1}(t)r(t)\nu(dt) \geqslant$$
$$\geqslant [r(d(c)) - r(c + \Delta)]\lambda \geqslant 0,$$

donde λ es el valor general de la integral (16).

Ahora nos convencemos de que $\psi(-\infty) < \varepsilon$. Designemos por t_0 la solución de la ecuación $r(t) = 1$. Si $d(-\infty) \leqslant t_0$, entonces $r(t) < 1$ en el intervalo $(-\infty, d(-\infty))$, y la igualdad requerida es, en virtud de (15), evidente. Si $d(-\infty) > t_0$, entonces, de un modo análogo obtenemos

$$\psi(-\infty) = 1 - \mathbf{P}_{\theta_1}(T \in (d(-\infty), \infty)) <$$
$$< 1 - \mathbf{P}_{\theta_1}(T \in (d(-\infty), \infty)) = \mathbf{P}_{\theta_1}(T \in (-\infty, d(-\infty))) = \varepsilon.$$

Exactamente igual se establece que $\psi(c_+) > \varepsilon$. ◁

Observación 1. Le dejamos al lector que el mismo se convenza de que para $\theta_1 < \theta_2$ la afirmación del teorema 2 y todas las investigaciones realizadas serán válidas si sustituimos el intervalo (θ_1, θ_2) por el segmento $[\theta_1, \theta_2]$, o sea, si verificamos la hipótesis $H_1 = \{\theta \in [\theta_1, \theta_2]\}$ frente a $H_2 = \{\theta \in \notin [\theta_1, \theta_2]\}$.

Observación 2. La exigencia del carácter exponencial de la familia $\{\mathbf{P}_\theta\}$, como se deduce de la demostración del teorema, puede ser debilitada hasta la condición de convexidad de la relación

$$q\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} + (1-q)\frac{f_{\theta_2}(X)}{f_{\theta_0}(X)}$$

con respecto a cierta estadística T (compárese con (10) y (11)).

Observación 3. Prestemos atención una vez más en que si la hipótesis principal fuera $H_2 = \{\theta \in (\theta_1, \theta_2)\}$, y la alternativa $H_1 = \{\theta \notin (\theta_1, \theta_2)\}$, entonces, el c.u.m.p. no existiría, ya que en este caso, los criterios "unilaterales" que tienen la forma $T > c$ o $T < c$ para las alternativas $\theta > \theta_2$ y $\theta < \theta_1$, respectivamente, resultarían más potentes que el criterio de forma $T \notin (c_1, c_2)$. Por ejemplo, para las alternativas $\theta > \theta_2$ existirá el c.u.m.p. de forma $T > c$, y la condición $\pi \in K_\varepsilon$ conducirá a la única limitación $\mathbf{M}_{\theta_2}\pi \leqslant \varepsilon$ (véanse las observaciones al final del punto 2).

No obstante, resultará que si la clase K_ε se reduce un poco adicionalmente, procediendo de un modo natural (véanse los §§ 6 y 7), entonces el c.u.m.p. también existirá en este problema.

3. Otro enfoque de los problemas sujetos a examen. La esencia matemática de la afirmación principal del teorema 2, así como de los teoremas en los §4 1 y 2, es muy simple y merece la pena que hablemos de ella especialmente. Por ejemplo, en el teorema 2, la misma consiste en el siguiente problema variacional. En la clase de funciones π que satisfacen las condiciones

$$\int \pi(x) f_{\theta_i}(x) \mu^n(dx) = \varepsilon, \quad i = 1, 2$$

buscamos el elemento π^0 para el cual se maximiza

$$\int \pi(x) f_{\theta_0}(x) \mu^n(dx).$$

La siguiente afirmación suele llamarse *generalización del lema fundamental de Neumann — Pearson.*

Lema 2. *Sean* $f_1, \ldots, f_{m+1}$ *las funciones reales definidas en* $\mathscr{X}^n$ *e integrables respecto a la medida* μ^n. *Supongamos que las funciones críticas* π *son tales, que*

$$\int \pi(x) f_i(x) \mu^n(dx) = e_i, \quad i = 1, \ldots, m. \tag{17}$$

Entonces, el elemento π°, *en el que* $\int \pi(x) f_{m+1}(x) \mu^n(dx)$ *alcanza el máximo, tiene la forma*

$$\pi^\circ(x) = \begin{cases} 1, & si \quad f_{m+1}(x) > \sum_{i=1}^{m} k_i f_i(x), \\ 0, & si \quad f_{m+1}(x) < \sum_{i=1}^{m} k_i f_i(x), \end{cases}$$

donde $k_1, \ldots, k_m$ se determinan de las condiciones (17).

Demostración. Designemos $F_i(\pi) = \int \pi(x) f_i(x) \mu^n(dx)$, $i = 1, \ldots, m + 1$. El elemento π que satisface las condiciones $F_i(\pi) = e_i$, $i = 1, \ldots, m$, maximiza $F_{m+1}(\pi)$ si y sólo si maximiza $F_{m+1}(\pi) - \sum_{i=1}^{m} k_i F_i(\pi)$ para cualesquiera $k_1, \ldots, k_m$ (pues el valor de la suma aquí está registrado). Por consiguiente, es suficiente que π maximice

$$\int \left(f_{m+1}(x) - \sum_{i=1}^{m} k_i f_i(x) \right) \pi(x) \mu^n(dx).$$

Pero esta expresión se vuelve máxima si se supone que $\pi(x) = 1$ allí donde $f_{m+1}(x) - \sum_{i=1}^{m} k_i f_i(x) > 0$, y $\pi(x) = 0$ allí donde esta expresión es negativa. Las constantes k_i, de las cuales depende este π, así como los valores "libres" de π en el conjunto $\left\{ f_{m+1}(x) = \sum_{i=1}^{m} k_i f_i(x) \right\}$, deben escogerse de modo que se cumpla (17). ◁

4. Enfoque bayesiano y distribuciones a priori menos favorables al construir el c.m.p. y el c.u.m.p. El lema 2 aclara la esencia matemática de las construcciones que hemos realizado en este párrafo. En el apartado presente también se tratará de la esencia de estas investigaciones, pero desde un punto de vista algo diferente. El hecho consiste en que al demostrar el teorema 2 hemos utilizado, implícitamente, el enfoque relacionado con la construcción de los criterios minimax a base de los criterios bayesianos (compárese con el teorema 1.2). Este enfoque se examina más detalladamente en la exposición sucesiva. Aquí obtendremos una afirmación general, útil para construir el c.u.m.p. en el caso general, y explicaremos su relación con el enfoque minimax.

Supongamos que se verifica la hipótesis fundamental $H_1 = \{\theta \in \Theta_1\}$ frente a la alternativa simple $H_2 = \{\theta = \theta_2\}$, $\theta_2 \notin \Theta_1$. En calidad de H_2 aquí también se puede tomar la alternativa arbitraria $\{X \in G\}$, donde G tiene una densidad g respecto a μ y no está de ningún modo relacionada con la familia $\{\mathbf{P}_\theta\}$. El problema consiste en determinar el c.m.p. $\bar{\pi}$ de nivel $1 - \varepsilon$ para verificar H_1 frente a H_2. Con otras palabras, es necesario hallar la función π de K_ε,

$$K_\varepsilon = \{\pi: \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi(X) \leqslant \varepsilon\} \tag{18}$$

que minimiza $\beta(\theta_2) = \mathbf{M}_{\theta_2}\pi(X)$. En las investigaciones precedentes hemos observado varias veces cierta dualidad en el planteamiento del problema: la maximización de la potencia, al ser registrada la probabilidad del error de primer género, equivale a la minimización de este último al ser registrada la potencia. Pero con tal inversión llegamos, en nuestra tarea, a la cuestión de minimización (18), que es precisamente el problema de construcción del criterio minimax (este problema se examina más detalladamente en el § 9). Ello explica, en cierta medida, la semejanza de la afirmación (que se demostrará más abajo) con el teorema 1.2.

Así pues, examinemos el planteamiento parcialmente bayesiano del problema, en virtud del cual el parámetro θ en el conjunto Θ_1 se elige al azar, con una distribución $\mathbf{Q}_1$. En este caso, la hipótesis compuesta H_1 se sustituye por la hipótesis simple H_{Q_1}, según la cual la densidad de X se define como el valor promediado respecto a la medida $\mathbf{Q}_1$:

$$f_{Q_1}(x) = \int_{\Theta_1} f_\theta(x)\mathbf{Q}_1(d\theta).$$

Para verificar H_{Q_1} frente a H_2 en la clase $K_\varepsilon^{Q_1} = \{\pi: \mathbf{M}_{Q_1}\pi(X) \leqslant \varepsilon\}$ de los criterios de nivel $1 - \varepsilon$ existe el c.m.p. π_{Q_1} que tiene la forma (π_{Q_1} es el criterio π_{Q_1,Q_2} en las designaciones del § 4, donde $\mathbf{Q}_2$ es la distribución degenerada en el punto θ_2):

$$\pi_{Q_1}(x) = \begin{cases} 1, & \text{si} \quad g(x) > cf_{Q_1}(x), \\ 0, & \text{si} \quad g(x) < cf_{Q_1}(x) \end{cases} \tag{19}$$

(aquí $g(x) = f_{\theta_2}(x)$ en el caso paramétrico).

Teorema 3. *Supongamos que existe tal distribución* $\mathbf{Q}_1$, *concentrada en el subconjunto* $\Theta_1^0 \subset \Theta_1(\mathbf{Q}_1(\Theta_1^0) = 1)$, *para la cual*

1) $$\pi_{Q_1} \in K_\varepsilon^{Q_1} \tag{20}$$

2) $$\mathbf{M}_\theta \pi_{Q_1}(X) = \text{const} = \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi_{Q_1}(X) \tag{21}$$

para todos $\theta \in \Theta_1^0$.

Entonces el criterio $\pi_{Q_1} \in K_\varepsilon$ *es precisamente el c.m.p. para la verificación de* H_1 *frente a* H_2.

Demostración. Comprobemos primeramente la pertenencia de $\pi_{Q_1} \in K_\varepsilon$. En virtud de las condiciones del teorema,

$$\sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi_{Q_1}(X) = \int_{\Theta_1^\varepsilon} \mathbf{M}_\theta \pi_{Q_1}(X) \mathbf{Q}_1(d\theta) = \mathbf{M}_{Q_1} \pi_{Q_1}(X) \leqslant \varepsilon. \quad (22)$$

Sea ahora π cualquier otro criterio de K_ε, o sea, el criterio de nivel $1 - \varepsilon$ para verificar H_1 frente a H_2. Entonces

$$\mathbf{M}_{Q_1} \pi(X) = \int \pi(x) f_{Q_1}(x) \mu^n(dx) = \int_{\Theta_1} \mathbf{M}_\theta \pi(X) \mathbf{Q}_1(d\theta) \leqslant \varepsilon$$

y, por lo tanto, $\pi \in K_\varepsilon^{Q_1}$. Pero entonces, en virtud de la definición de π_{Q_1},

$$\mathbf{M}_{\theta_2} \pi_{Q_1}(X) \geqslant \mathbf{M}_{\theta_2} \pi(X),$$

que es lo que se necesitaba demostrar. ◁

La distribución $\mathbf{Q}_1$ que figura en el teorema se llama *distribución menos favorable.* Esto está relacionado con la circunstancia siguiente. La magnitud $\beta_{Q_1}(\theta_2) = \mathbf{M}_{\theta_2} \pi_{Q_1}(X)$ es el mayor valor de potencia que puede ser alcanzado *en $K_\varepsilon^{Q_1}$ con la distribución "a priori" Q_1 en Θ_1. Si tomamos* ahora cualquier otra distribución Q' en Θ_1, obtenemos

$$\beta_{Q'}(\theta_2) \geqslant \beta_{Q_1}(\theta_2), \quad \beta_{Q_1}(\theta_2) = \inf_{Q'} \beta_{Q'}(\theta_2)$$

(esto es precisamente el sentido del término "la peor distribución"). En efecto, en virtud de (22) π_{Q_1} pertenece a K_ε y, por lo tanto, a $K_\varepsilon^{Q'}$. Esto quiere decir que su potencia $\beta_{Q_1}(\theta_2) = \mathbf{M}_{\theta_2} \pi_{Q_1}(X)$ no superará la potencia del c.m.p. en $K_\varepsilon^{Q'}$ que, por definición, es igual a $\beta_{Q'}(\theta_2)$.

Ahora, con ayuda del teorema 3 podríamos demostrar los teoremas 1 y 2. El conjunto Θ_1^ε, en el que está concentrada la distribución menos favorable, en los teoremas 1 y 2 consta de un solo $\{\theta_1\}$ y de los puntos $\{\theta_1, \theta_2\}$, respectivamente. Las condiciones (20) y (21) se transforman, respectivamente, en condiciones (3) y (7).

Análogamente ha de utilizarse el teorema 3 para construir el c.u.m.p. en otros casos: si el criterio construido π_{Q_1} no depende de $\theta_2 \in \Theta_2$, entonces él será el c.u.m.p. para verificar $H_1 = \{\theta \in \Theta_1\}$ frente a $H_2 = \{\theta \in \Theta_2\}$ en la clase K_ε.

La distribución menos favorable $\mathbf{Q}_1$, que satisface las condiciones del teorema 3, existe para suposiciones muy amplias que suelen cumplirse en los problemas reales. Es suficiente exigir la compacticidad de Θ_1 y la continuidad de $f_\theta(x)$ respecto a θ para x c.d. (véase [57] y los capítulos posteriores).

La investigación ulterior de las relaciones entre los enfoques bayesiano y minimax véase en el § 9.

§ 6*. Criterios no desplazados

En este párrafo y en el siguiente utilizaremos los principios de no desplazamiento y de invariación para la reducción natural de la clase de criterios que se examinan. El objetivo de tal reducción consiste en determinar los criterios óptimos.

1. Definiciones y c.u.m.p. no desplazados. Al igual que en el párrafo anterior, examinaremos la verificación de la hipótesis compuesta $H_1 = \{\theta \in \Theta_1\}$ frente a $H_2 = \{\theta \in \Theta_2\}$, basándonos en la muestra $X \in \mathbf{P}_\theta$, $\theta \in \Theta = \Theta_1 \cup \Theta_2$. Examinemos primero los criterios π de la clase $K_\varepsilon = \{\pi: \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi \leqslant \varepsilon\}$.

Si, por ejemplo, Θ_1 comprende un solo punto θ_1, $\mathbf{M}_{\theta_1}\pi = \varepsilon$, entonces ε es la probabilidad de que se rechace H_1 cuando H_1 es cierta. La exigencia natural respecto al criterio π consiste en que la *probabilidad de rechazar H_1, cuando H_1 no es cierta, ha de ser mayor que ε*. Si no es así, entonces habrá alternativas con las que la aceptación de H_1 será más probable que en los casos cuando H_1 es cierta. Tal situación es indeseable. Llegamos a la necesidad de destacar la siguiente clase importante de criterios.

Definición 1. El criterio π se llama *no desplazado* si para él

$$\inf_{\theta \in \Theta_2} \mathbf{M}_\theta \pi(x) \geqslant \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi(X). \tag{1}$$

Ahora bien, el criterio $\pi \in K_\varepsilon$ (para el cual $\sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi = \varepsilon$) no estaría desplazado si $\beta_\pi(\theta) \geqslant \varepsilon$ cuando $\theta \in \Theta_2$. La clase de criterios no desplazados de nivel $1 - \varepsilon$ se designa por $\bar{K}_\varepsilon$.

El criterio unilateral π con región crítica $T > c$ (o $T < c$) para familias exponenciales, mencionado en el párrafo anterior, no puede permanecer sin desplazamiento al verificar $H_1 = \{X \in \mathbf{P}_{\theta_1}\}$ frente a $H_2 = \{X \in \mathbf{P}_\theta, \theta \neq \theta_1\}$, ya que aquí $\Theta_2 = \{\theta: \theta \neq \theta_1\}$, $\mathbf{M}_\theta \pi < \varepsilon$ para $\theta < \Theta_1$ si $\mathbf{M}_{\theta_1}\pi = \varepsilon$ (véase el teorema 5.1).

Al contrario, los c.u.m.p., si existen, con la necesidad pueden no estar desplazados, ya que para ellos la potencia $\beta(\theta)$, cuando $\theta \in \Theta_2$, no puede ser menor que la potencia del criterio $\pi(X) \equiv \varepsilon$.

El principio de no desplazamiento*) reviste interés especial, puesto que permite reducir naturalmente la clase de criterios. Esto nos permite construir los c.u.m.p. en las clases $\bar{K}_\varepsilon$ cuando los c.u.m.p. no existen en la clase K_ε.

---

*) El término "no desplazamiento" también se utilizó con arreglo a las estimaciones. Desde cierto punto de vista la propiedad de no desplazamiento de la estimación es análoga a la propiedad de no desplazamiento del criterio: si la estimación θ^* no está desplazada, entonces $\mathbf{M}_{\theta_0}\theta^* \neq \theta_0$ y habrá otros valores del parámetro $\theta \neq \theta_0$ con los cuales el valor medio $\mathbf{M}_\theta \theta^*$ será igual a θ_0.

Como veremos, esto se refiere, en particular, al problema de verificación de la hipótesis $H_1 = \{\theta \in [\theta_1, \theta_2]\}$, $\theta_1 \leqslant \theta_2$, frente a la alternativa bilateral $H_2 = \{\theta \notin [\theta_1, \theta_2]\}$ (compárese con el apartado 2 del § 5).

La determinación de los criterios no desplazados y uniformemente más potentes puede ser bastante reducida al uso de los procedimientos ya empleados, cuya esencia se expone en el lema 5.2. En este caso puede ser útil la siguiente afirmación.

Supongamos que existe una frontera común no vacía Γ de los conjuntos Θ_1 y Θ_2 de R^k:

$$\Gamma = \partial\Theta_1 \cap \partial\Theta_2$$

($\partial\Theta_i$ designa la frontera de Θ_i), o sea, un conjunto de puntos límites para Θ_1 y Θ_2. Supongamos además, que para todos $\pi \in K_\varepsilon$,

$$\beta_\pi(\theta) = \mathbf{M}_\theta \pi(X) = \varepsilon \quad \text{cuando} \quad \text{todos } \theta \in \Gamma. \tag{2}$$

Es evidente que esta propiedad siempre se cumplirá si $\beta_\pi(\theta)$ depende continuamente de θ para cualquier criterio π de K_ε.

Como

$$\beta_\pi(\theta) = \int \pi(x) f_\theta(x) \mu^n(dx), \quad 0 \leqslant \pi(x) \leqslant 1,$$

entonces la continuidad de $\beta_\pi(\theta)$ tendrá lugar si la función $f_\theta(x)$ es continua respecto a θ para c.t. μ^n de x. Esto se deduce del corolario 1 del Suplemento VI.

Designemos por $\bar{K}_\varepsilon$ la clase de todos los criterios π que satisfacen (2).

Lema 1. *Supongamos que* $K_\varepsilon \subset \bar{K}_\varepsilon$ *(o sea, que se cumple* (2)*). En este caso, si* $\check{\pi}$ *es el c.u.m.p. en* $\bar{K}_\varepsilon \cap K_\varepsilon$, *entonces* $\check{\pi}$ *es el c.u.m.p. en* K_ε.

Demostración. Es suficiente convencerse que $\check{\pi} \in K_\varepsilon$ y que $K_\varepsilon \subset \subset \bar{K}_\varepsilon \cap K_\varepsilon$. La segunda de estas relaciones se desprende de la suposición de que $K_\varepsilon \subset \bar{K}_\varepsilon$. La primera se deduce del hecho de que el criterio $\pi \equiv \varepsilon$ partenece a $\bar{K}_\varepsilon \cap K_\varepsilon$ y, por lo tanto, $\inf\limits_{\theta \in \Theta_2} \mathbf{M}_\theta \check{\pi}(X) \geqslant \inf\limits_{\theta \in \Theta_2} \mathbf{M}_\theta \pi = \varepsilon$. ◁

Ahora bien, el lema 1 permite reducir la búsqueda de los criterios

ciones (2) serán dos ecuaciones $\mathbf{M}_{\theta_i}\pi(X) = \varepsilon$, $i = 1, 2$. Sin embargo, en el caso límite $\theta_1 = \theta_2$, estas ecuaciones se transforman en una sola. Pero en virtud del no desplazamiento del criterio π, su potencia $\beta_\pi(\theta)$ debe alcanzar su mínimo en el punto θ_1 (véase (1)). Por consiguiente, si $\beta_\pi(\theta)$ es derivable, entonces, el papel de las ecuaciones (2) en el caso de las $\theta_1 = \theta_2$ lo desempeñarán las igualdades

$$\beta_\pi(\theta_1) = \varepsilon, \quad \beta_\pi'(\theta_1) = 0. \tag{3}$$

Las condiciones de derivabilidad de $\int f_\theta(x)\mu(dx)$ y, por consiguiente, también de $\beta_\pi(\theta) = \mathbf{M}_\theta\pi(X)$, son aclaradas en el Suplemento VI. Si se cumplen estas condiciones, entonces

$$\beta_\pi'(\theta) = \int \pi(x) f_\theta'(x)\mu^n(dx) =$$
$$= \int \pi(x) L'(x, \theta) f_\theta(x)\mu^n(dx) = \mathbf{M}_\theta\pi(X)L'(X, \theta).$$

Esto significa que las condiciones (3) pueden escribirse de nuevo en términos integrales:

$$\mathbf{M}_{\theta_1}\pi(X) = \varepsilon, \quad \mathbf{M}_{\theta_1}\pi(X)L'(x, \theta_1) = 0. \tag{4}$$

Por ejemplo, para la familia exponencial (5.9),

$$L'(x, \theta) = c'(\theta)/c(\theta) + a'(\theta)T(x).$$

Como $\mathbf{M}_\theta L'(x, \theta) = 0$, entonces $c'(\theta)/c(\theta) = -a'(\theta)\mathbf{M}_\theta T(X)$,

$$\mathbf{M}_\theta\pi(X)L'(X, \theta) = -a'(\theta)\mathbf{M}_\theta T(X)\cdot\mathbf{M}_\theta\pi(x) + a'(\theta)\mathbf{M}_\theta\pi(X)T(X),$$

y las ecuaciones (4) adoptan la forma

$$\mathbf{M}_{\theta_1}(\pi(X) - \varepsilon) = 0, \quad \mathbf{M}_{\theta_1}(\pi(X) - \varepsilon)T(X) = 0.$$

En calidad de ejemplo ilustremos un caso para cuyo examen, de hecho, ya todo está preparado.

2. Alternativas bilaterales. Familia exponencial.

Teorema 1. *Supongamos que $f_\theta(x)$ se define por la igualdad (5.9), y que se verifica la hipótesis $H_1 = \{\theta \in [\theta_1, \theta_2]\}$, $\theta_1 \leqslant \theta_2$, frente a la alternativa $H_2 = \{\theta \notin [\theta_1, \theta_2]\}$. Entonces, si la función $a(\theta)$ es monótona,*

1) *en la clase K_ε de criterios no desplazados de nivel $1 - \varepsilon$ existe un c.u.m.p. $\breve{\pi}$ que tiene la forma siguiente:*

$$\breve{\pi}(x) = \begin{cases} 0 & si \quad c_1 < T(x) < c_2, \\ p_i & si \quad T(x) = c_i, \ i = 1, 2, \\ 1 & si \quad T(x) \notin [c_1, c_2], \end{cases} \tag{5}$$

donde $T(x) = \sum_{i=1}^{n} U(x_i)$, y las constantes c_i, p_i, $i = 1, 2$ se deducen de las

condiciones

$$\mathbf{M}_{\theta_i}\check{\pi}(X) = \varepsilon, \quad i = 1, 2, \tag{6}$$

si $\theta_1 < \theta_2$, *y de las condiciones*

$$\mathbf{M}_{\theta_1}\check{\pi}(X) = \varepsilon, \quad \mathbf{M}_{\theta_1}(\check{\pi}(X) - \varepsilon)T(X) = 0, \tag{7}$$

si $\theta_1 = \theta_2$.

2) *El criterio* $\check{\pi}$ *minimiza la función* $\beta_\pi(\theta) = \mathbf{M}_\theta\pi(X)$ *en las condiciones* (6) *dentro del segmento* $[\theta_1, \theta_2]$, *y la maximiza fuera de* $[\theta_1, \theta_2]$ *en las condiciones* (6) ó (7) (*esto último sucede cuando* $\theta_1 = \theta_2$).

3) *cuando* $0 < \varepsilon < 1$ y $\theta_1 < \theta_2$, *la función* $\check{\beta}(\theta) = \mathbf{M}_\theta\check{\pi}(X)$ *alcanza su valor mínimo en cierto punto* $\theta_0 \in (\theta_1, \theta_2)$ y crece estrictamente al alejarse θ de θ_0 a la derecha o a la izquierda. *Además, excluimos el caso* (5.8).

No es difícil ver que la enunciación de este teorema casi repite la afirmación del teorema 5.2. La única diferencia consistente en que las propias afirmaciones tienen, a veces, carácter "contrario" y no se excluye la igualdad $\theta_1 = \theta_2$.

Demostración. En el caso de $\theta_1 < \theta_2$, ésta es absolutamente análoga a la demostración del teorema 5.2. En la nota 1 adjunta a este teorema hemos dicho que para $\theta_1 < \theta_2$ todos los razonamientos del referido teorema conservan su validez en el caso cuando se verifica la hipótesis $\{\theta \not\subset [\theta_1, \theta_2]\}$ frente a $\{\theta \in [\theta_1, \theta_2]\}$, o sea, a los símbolos de este párrafo: la hipótesis H_2 frente a la H_1. Pongamos $\check{\pi}(x) = 1 - \pi^\circ(x)$, donde π° es la función definida en (5.6) para las condiciones $\mathbf{M}_{\theta_i}\pi^\circ(X) = 1 - \varepsilon$, $i = 1, 2$, en vez de (5.7). Entonces, las afirmaciones 2) y 3) serán, evidentemente, los corolarios directos de las respectivas afirmaciones del teorema 5.2.

La primera afirmación del teorema resulta de la segunda, ya que la clase de criterios π que satisfacen (6) es más amplia que $\check{K}_\varepsilon$ y, por consiguiente, $\check{\pi}$ maximizará $\mathbf{M}_\theta\pi(x)$ en la clase $\check{K}_\varepsilon$ en cualquier punto θ fuera de $[\theta_1, \theta_2]$. Esto significa que π es el criterio no desplazado uniformemente más potente.

Nos queda examinar el caso $\theta_1 = \theta_2$. Aquí es más simple, por lo visto, hacer uso del lema 5.2. Tomemos cualquier $\theta \neq \theta_1$ y examinemos el problema de maximización de $\mathbf{M}_\theta\pi(X)$ para las condiciones

$$\mathbf{M}_{\theta_1}\pi(x) = \varepsilon, \quad \mathbf{M}_{\theta_1}\pi(X)T(X) = \varepsilon\mathbf{M}_{\theta_1}T(X).$$

Es evidente que nos encontraremos en condiciones del lema 5.2 si ponemos $m = 2$, $f_1 = f_{\theta_1}$, $f_2 = Tf_{\theta_1}$, $f_3 = f_\theta$, $\varepsilon_1 = \varepsilon$, $\varepsilon_2 = \varepsilon\mathbf{M}_\theta T(X)$. Según este lema, el máximo $\mathbf{M}_\theta\pi$ se alcanzará en la función

$$\check{\pi}(x) = \begin{cases} 1, & \text{si} \quad f_\theta(x) > k_1 f_{\theta_1}(x) + k_2 T(x) f_{\theta_1}(x), \\ 0, & \text{si} \quad f_\theta(x) < k_1 f_{\theta_1}(x) + k_2 T(x) f_{\theta_1}(x). \end{cases}$$

Examinemos la última desigualdad, que puede ser escrita en la forma

$$\frac{c(\theta)}{c(\theta_1)} e^{(a(\theta) - a(\theta_1))T(x)} < k_1 + k_2 T(x).$$

Está claro que para todos $c_1 < c_2$ siempre se puede escoger k_1, k_2 de modo que esta desigualdad equivalga a

$$c_1 < T < c_2.$$

Esto demuestra que el criterio de forma (5) maximiza $\mathbf{M}_\theta \pi(X)$ en las condiciones (7) siempre que c_i y p_i, $i = 1, 2$ puedan escogerse en (5) de modo que se satisfaga (7) (u (8)). Este criterio será, evidentemente, el criterio no desplazado uniformemente más potente, ya que la clase de criterios π que satisfacen (8) es más amplia que $\tilde{K}_\varepsilon$ y, por lo tanto, $\check{\pi}$ también maximizará $\mathbf{M}_\theta \pi(X)$ en $\tilde{K}_\varepsilon$. Así pues, para demostrar el teorema queda demostrar que es válido el

Lema 2. *La ecuación* (7) *cuando* $0 < \varepsilon < 1$ *es resoluble respecto a* c_i *y* p_i, $i = 1, 2$.

La demostración de este lema, al igual que la del lema 5.1, será expuesta suponiendo simplemente *que la* $\mathbf{P}_{\theta_1}$-distribución de $T(X)$ es continua, es decir, $\mathbf{P}_{\theta_1}(T(X) = c) = 0$ para todos c.

Recordemos que la densidad de la distribución T respecto a cierta medida ν puede considerarse igual a (véase el § 5) $g_\theta(t) = c(\theta)e^{\theta t}$. Entonces, las ecuaciones (7) y (8) serán equivalentes a las relaciones

$$\mathbf{M}_{\theta_1}(1 - \pi(X)) = c(\theta_1) \int_{c_1}^{c_2} e^{\theta_1 t}\nu(dt) = 1 - \varepsilon, \tag{9}$$

$$\mathbf{M}_{\theta_1}(1 - \pi(X))T(X) = c(\theta_1) \int_{c_1}^{c_2} te^{\theta_1 t}\nu(dt) = (1 - \varepsilon)c(\theta_1) \int_{c_1}^{c_2} te^{\theta_1 t}\nu(dt).$$

Designando $r(t) = t$, $m = \mathbf{M}_{\theta_1} T(X) = c(\theta_1)\int te^{\theta_1 t}\nu(dt)$, podemos escribir las ecuaciones (9) en la forma

$$c(\theta_1) \int_{c_1}^{c_2} e^{\theta_1 t}\nu(dt) = 1 - \varepsilon,$$

$$c(\theta_1) \int_{c_1}^{c_2} r(t)e^{\theta_1 t}\nu(dt) = (1 - \varepsilon)m. \tag{10}$$

Hemos llegado al problema que coincide con el problema examinado en el lema 5.1. La única diferencia consistente en que la distribución con densidad $r(t)g_{\theta_1}(t)$ puede ser generalizada (o sea, también puede adoptar valores negativos). En estas nuevas condiciones conviene poner $t_0 = m$. En lo demás, los razonamientos del lema 5.1 no cambian. ◁

§ 7*. Criterios invariantes.

En este párrafo examinaremos otra manera de reducir la clase de todos los criterios, basada, esta vez, en las consideraciones de invariación.

Supongamos que $X \in \{\mathbf{P}_\theta\}$ y que $\{\mathbf{P}_\theta\}$ es una familia invariante. Recordemos las designaciones necesarias y los conceptos respectivos (véase el § 2.19). Supongamos asimismo, que se ha dado un grupo G de transformaciones medibles g del espacio $\mathscr{X}^n$ en sí. La familia $\{\mathbf{P}_\theta\}$ será invariante respecto a G, si para cada $g \in G$ y cada $\theta \in \Theta$ hay un elemento $\theta_g \in \Theta$ tal, que

$$\mathbf{P}_{\theta_g}(X \in A) = \mathbf{P}_\theta(gX \in A)$$

para cualquier $A \in \mathfrak{B}^n_{\mathscr{X}}$.

Las transformaciones $\bar{g}$ del espacio Θ, definidas por la igualdad $\bar{g}\theta = \theta_g$, forman, al cumplirse las condiciones A_0, el grupo $\bar{G}$ (véase el § 2.19).

Definición 1. Diremos que el problema de verificación de la hipótesis $H_1 = \{\theta \in \Theta_1\}$ frente a $H_2 = \{\theta \in \Theta_2\}$, $\Theta_1 \cup \Theta_2 = \Theta$ es invariante siempre que se cumplan las dos condiciones siguientes:

1) La familia $\{\mathbf{P}_\theta\}$ es invariante respecto a G.

2) Los conjuntos Θ_1 y Θ_2 son invariantes respecto a $\bar{g} \in \bar{G}$, o sea, $\bar{g}\Theta_i = \Theta_i$, $i = 1, 2$.

Si el problema de verificación de las hipótesis es invariante, es natural que para su solución se haga uso del criterio invariante.

Definición 2. El criterio π se llama *invariante* cuando $\pi(x)$ es estadística invariante respecto a g*):

$$\pi(gx) = \pi(x) \quad \text{para todos } x \in \mathscr{X}^n, \quad g \in G.$$

Si π es un criterio no randomizado y Ω_j es la región de aceptación de la hipótesis H_j, entonces, la invariación de π significará que $g\Omega_j = \Omega_j$, $j = 1, 2$.

La utilización natural de los criterios invariantes se puede comprender, por lo visto, con más facilidad, a base de ejemplos. La investigación general, relacionada con la interpretación de g como la sustitución de las coordenadas y la insensibilidad de las estadísticas respectivas a esta sustitución, está contenida en el § 2.19.

Ejemplo 1. Los ejemplos más simples se refieren al caso cuando el grupo $\bar{G}$ es trivial, o sea, cuando $\bar{g}$ para todo g es la transformación idéntica $\bar{e}$ del espacio Θ.

Supongamos que $X \in \Phi_{0,\sigma^2}$; se verifica la hipótesis $H_1 = \{\sigma_1 \leqslant \sigma \leqslant \sigma_2\}$ frente a la alternativa adicional H_2. En este caso

$$f_{\sigma^2}(X) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right\}.$$

*) Véase la nota en la pág. 195.

Es evidente que la familia Φ_{0,σ^2} es invariante respecto al grupo G de transformaciones ortogonales g (revoluciones) del espacio $\mathscr{X}^n$, con la particularidad de que $\bar{g} = \bar{e}$ para cualquier g. Por eso es natural examinar los criterios π que dependen exclusivamente de la estadística $T(X) = \sum_{i=1}^{n} x_i^2$. En vista de que $\sigma^{-2}T(X) \in \Gamma_{1/2,n/2} = H_n$, entonces $T(X) \in \Gamma_{\alpha,n/2}$ para $\alpha = 1/(2\sigma^2)$ y llegamos al problema de verificación de la hipótesis $H_1 = \{\alpha_1 \leqslant \alpha \leqslant \alpha_2\}$, $\alpha_1 = 1/(2\sigma_2^2)$, $\alpha_2 = 1/(2\sigma_1^2)$, según la observación $T(X)$ que tiene la distribución $\Gamma_{\alpha,n/2}$ de una familia exponencial. Con ayuda de los resultados de los párrafos precedentes podemos construir el criterio no desplazado y uniformemente más potente, de nivel $1 - \varepsilon$, que acepta H_1 cuando

$$c_1 \leqslant T(X) \leqslant c_2, \tag{1}$$

donde c_i se elige de modo que $\Gamma_{\alpha_1,n/2}(R \setminus [c_1, c_2]) = \Gamma_{\alpha_2,n/2}(R \setminus [c_1, c_2]) = \varepsilon$.

Nótese que en este ejemplo podríamos construir el criterio de la forma (1) partiendo también de otras consideraciones, o sea, basándonos en el principio de insuficiencia, ya que la estadística T es suficiente. Pues sabemos que toda la información acerca del parámetro σ^2 está concentrada en T y no vale la pena utilizar otras estadísticas (o sea, otra información relacionada con la muestra).

En lo sucesivo, allí donde sea posible, reduciremos inmediatamente este problema al problema de distribución de las estadísticas suficientes.

Ejemplo 2. Supongamos que $X \in \Phi_{\alpha,\sigma^2}$, $H_1 = \{\sigma_1 \leqslant \sigma \leqslant \sigma_2\}$. En este caso $\theta = (\alpha, \sigma^2)$ y la transformación de desplazamiento $gX = X + c = (x_1 + c, \ldots, x_n + c)$ induce la transformación $\bar{g}\alpha = \alpha + c$ que mantiene invariable la hipótesis H_1. Si nos limitamos a investigar las estadísticas suficientes

$$T_1 = \bar{x}, \quad T_2 = \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

entonces, la transformación g proporcionará

$$T_1(gX) = \bar{x} + c, \quad T_2(gT) = T_2(X).$$

Ahora bien,, la estadística T_2 es invariante respecto a G. Es decir, el criterio invariante π, basado en las estadísticas suficientes, debe ser una función de T_2. (Más adelante veremos que *cualquier* criterio invariante π debe ser una función de T_2). En virtud del § 2.32, $\sigma^{-2}T_2 \in \Gamma_{1/2,(n-1)/2}$ y llegamos al problema examinado en el ejemplo precedente. El criterio invariante no desplazado y uniformemente más potente tendrá la forma $c_1 \leqslant T_2 \leqslant c_2$.

Ejemplo 3. Los dos ejemplos examinados más arriba se referían a la distribución normal. Con arreglo a la distribución de la muestra X, la misma era una distribución normal multidimensional con una matriz diagonal de segundos momentos. Para la exposición posterior es útil notar que la familia de distribuciones normales multidimensionales arbitrarias Φ_{α,σ^2}, $\alpha \in R^m$, $\sigma^2 = \|\sigma_{ij}\|$, $i, j = 1, \ldots, m$ es invariante respecto al grupo G de transformaciones no degeneradas lineales

$$gx = (x - a)C,$$

donde C es una matriz inversa. En efecto, debemos convencernos que, con cierta transformación $\bar{g}$, se cumple $\mathbf{P}_{\bar{g}\theta}(A) = \mathbf{P}_\theta(g^{-1}A)$, donde $\mathbf{P}_\theta \equiv \Phi_{\alpha,\sigma^2}$, $\theta = (\alpha, \sigma^2)$, $g^{-1}A$ significa, por lo común, el conjunto $g^{-1}A = \{x \in R^m\colon gx \in A\}$. Tenemos ($\sigma = \sqrt{|\sigma^2|}$)

$$\Phi_{\alpha,\sigma^2}(g^{-1}A) = \frac{1}{(2\pi)^{m/2}\sigma} \int\limits_{g^{-1}A} \exp\left\{-\frac{1}{2}(x - \alpha)\sigma^{-2}(x - \alpha)^T\right\} dx.$$

Después de sustituir $y = gx$, obtenemos

$$\Phi_{\alpha,\sigma^2}(g^{-1})A = \frac{1}{(2\pi)^{m/2}\sigma|C|} \int\limits_{A} \exp\left\{-\frac{1}{2}(g^{-1}y - \alpha)\sigma^{-2}(g^{-1}y - \alpha)^T\right\} dy.$$

Teniendo en cuenta que $g^{-1}y = yC^{-1} + a$, en la última integral podemos escribir el exponente de la forma siguiente:

$$(y - (\alpha - a)C^{-1}\sigma^{-2}(C^{-1})^T(y - (\alpha - a)C)^T.$$

Por consiguiente, si se pone

$$\bar{g}\theta = \bar{g}(\alpha, \sigma^2) = (g\alpha, C^T\sigma^2C) = ((\alpha - a)C, C^T\sigma^2C), \tag{2}$$

obtenemos

$$\Phi_{\alpha,\sigma^2}(g^{-1}A) = \Phi_{\bar{g}(\alpha,\sigma^2)}(A). \tag{3}$$

Ejemplo 4. Supongamos que las hipótesis H_j tienen la forma siguiente: $H_j = \{X \in \mathbf{P}_{j,\alpha}\}$, $\alpha \in \mathscr{X}$, $j = 1, 2$, donde $\mathbf{P}_{j,\alpha}$ son las distribuciones con densidades $f_j(x - \alpha)$, $j = 1, 2$. Con otras palabras, nos interesa a cuál de dos tipos de distribuciones le pertenece, con una exactitud de hasta el desplazamiento, la muestra X. Aquí conviene poner $\theta = (\nu, \alpha)$, $\nu = 1, 2$, $\alpha \in \mathscr{X}$ y examinar la transformación $gX = X + c$ que en el espacio paramétrico induce la transformación $\bar{g}(\theta = (\nu, \alpha + c)$. Está claro que las hipótesis $H_j = \{\nu = j\}$, $j = 1, 2$ son invariantes respecto a $\bar{g}$ y, por lo tanto, el problema de verificación de estas hipótesis también es invariante. La estadística

$$Y = (x_1 - x_n, \ldots, x_{n-1} - x_n)$$

será invariante respecto a g (compárese con el § 2.18). La distribución de esta estadística en el punto $y = (y_1, \ldots, y_{n-1})$, en caso de la hipótesis H_j, tiene la densidad siguiente:

$$f_j^Y(y) = \int \prod_{i=1}^{n-1} f_j(y_i + z) f_j(z) \mu(dz). \tag{4}$$

De aquí se deduce que para la observación Y, las hipótesis H_j se transforman en hipótesis simples, conforme a las cuales las densidades f_j^Y para Y tienen la forma (4). En estas condiciones podemos hacer uso del lema de Neumann — Pearson y construir el c.m.p. π que acepta la hipótesis H_2 si

$$f_2^Y(Y)/f_1^Y(Y) > c. \tag{5}$$

Como este criterio no depende de α, el mismo será el c.u.m.p. para verificar H_1 frente a H_2 entre todos los criterios invariantes basados en la estadística Y.

Con arreglo a los ejemplos examinados es conveniente estar seguro de que los demás criterios invariantes en estos problemas también son funciones de las estadísticas invariantes escogidas por nosotros. Esto se refiere especialmente al último ejemplo, puesto que en los dos ejemplos anteriores, la elección de los criterios también se basaba en las consideraciones de suficiencia.

Para aclarar las relaciones mutuas entre los invariantes, introduzcamos algunos conceptos. Dos puntos x y x' de $\mathscr{X}^n$ se llamarán *equivalentes* respecto al grupo G si existe $g \in G$ tal, que $x' = gx$. Como G es un grupo, entonces todo el espacio $\mathscr{X}^n$ se divide en clases disjuntas de equivalencia, que en el § 2.19 hemos llamado *órbitas*. Para obtener cierta órbita es suficiente tomar un punto cualquiera x_0 de la misma y aplicar a éste todas las transformaciones g de G. Por ejemplo, para las transformaciones ortogonales del ejemplo 1, las órbitas forman esferas cuyos centros coinciden con el origen de coordenadas.

La invariación de la estadística T respecto a G es unívoca al hecho de que T es constante en cada órbita.

Definición 3. La estadística T se denomina *invariante máximo* si la misma es invariante, y de $T(x') = T(x)$ se deduce $x' = gx$ para cierto $g \in G$.

Esto significa que el invariante máximo adopta distintos valores en órbitas diferentes.

Teorema 1. *Sea T el invariante máximo. La estadística S es invariante si y sólo si S depende de X a través de t, o sea,* si existe una función φ *tal, que* $S(X) = \varphi(T(X))$.

Para simplificar la exposición, aquí no tratamos una cuestión importante, relacionada con la mensurabilidad de φ. Nótese solamente que en los ejemplos examinados en este párrafo, tal mensurabilidad tendrá lugar[*)].

Demostración. Si $S(x) = \varphi(t(x))$, entonces $S(gx) = \varphi T(gx) = \varphi(T(x)) = S(x)$ y, por lo tanto, S es invariante. Para demostrar la afirmación inversa debemos convencernos de que de $T(x) = T(x')$ resulta $S(x) = S(x')$. Pero esto es así en virtud del hecho de que $T(x) = T(x')$ provoca la existencia de una g tal, que $x' = gx$. Pero como S es un invariante, $S(x) = S(x')$. ◁

A título de ejemplo examinemos el grupo G de desplazamientos

$$gx = x + c = (x_1 + c, \ \dots, x_n + c).$$

Como ya hemos señalado, la estadística $Y(x) = (x_1 - x_n, \dots, x_{n-1} - x_n)$ es un invariante. Mostremos que éste es el invariante máximo. En efecto, de $Y(x) = Y(x') = (x_1' - x_n', \dots, x_{n-1}' - x_n')$ se desprende que $x_i - x_n = x_i' - x_n'$ para todos $i = 1, \dots, n - 1$. Poniendo $x_n' - x_n = c$, obtenemos $x_i' = x_i + c$, $i = 1, \dots, n$, $x' = x + c = gx$, lo que precisamente significa la equivalencia necesaria de x' y x.

Ahora podemos volver al ejemplo 3 y afirmar que el criterio (5) es el c.u.m.p. entre todos los criterios invariantes, puesto que según el teorema 1 todos los criterios invariantes son funciones de Y y, por consiguiente, la suposición de que exista un criterio invariante más potente que (5) será contradictoria.

Por analogía a lo expuesto anteriormente, el lector puede convencerse de que la estadística $\sum_{i=1}^{n} x_i^2$ en el ejemplo 1 también es un invariante máximo.

Si existen estadísticas suficientes, al principio suele ser conveniente reducir el problema inicial al problema respecto a la distribución de las estadísticas suficientes y luego emplear las consideraciones de invariación así como se hizo en el ejemplo 2, donde la estadística $T_2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$ es, evidentemente, el máximo invariante en la observación $(\bar{x}, T_2)$.

En conclusión de este párrafo es preciso señalar una vez más, que la esencia del enfoque relacionado con la invariación consiste en que los problemas sometidos a examen y destinados a la verificación de las hipótesis, deben reducirse a problemas más simples, referentes a la distribución de los invariantes máximos. En estas nuevas condiciones, que son más simples, resulta posible, en varios casos, construir el c.m.p. o el c.u.m.p.

---

*) Véanse, por ejemplo, [57] y [95].

En este sentido, el "principio de invariación" se asemeja a los "principios" de suficiencia y de no desplazamiento, de acuerdo con los cuales el problema inicial se reduce a un problema en términos de estadística suficiente o de estadística no desplazada.

§ 8*. Enlace con los conjuntos confidenciales.

1. Enlace de los criterios estadísticos y los conjuntos confidenciales. Enlace de las propiedades de optimización. Los conceptos de conjunto confidencial y de criterio estadístico están estrechamente ligados entre sí. En el § 2.31 hemos dado la definición del conjunto confidencial. Recordémosla.

Sea $X \in \mathbf{P}_\theta$, $\theta \in \Theta$.

Definición 1. El subconjunto aleatorio $\Theta^* = \Theta^*(x, \varepsilon)$ del espacio estadístico Θ se llama *conjunto confidencial de nivel* $1 - \varepsilon$, si

$$\mathbf{P}_\theta(\Theta^*(X, \varepsilon) \ni \theta) \geqslant 1 - \varepsilon \tag{1}$$

para todos $\theta \in \Theta$.

Evidentemente, el intervalo confidencial es un caso particular del conjunto confidencial. Este último tiene el mismo sentido: con una probabilidad $\geqslant 1 - \varepsilon$ recubre el valor verdadero del parámetro.

Designemos

$$\Omega(\theta, \varepsilon) = \{x \in \mathscr{X}^n: \theta \in \Theta^*(x, \varepsilon)\}. \tag{2}$$

Entonces, las relaciones

$$\theta \in \Theta^*(x, \varepsilon) \quad \text{y} \quad x \in \Omega(\theta, \varepsilon) \tag{3}$$

serán equivalentes.

La definición del conjunto confidencial supone que el conjunto $\Omega(\theta, \varepsilon)$ en (2) es medible, así que la probabilidad en (1) tiene sentido y es igual a $\mathbf{P}_\theta(X \in \Omega(\theta, \varepsilon))$.

Los conjuntos confidenciales y los criterios estadísticos para verificar la hipótesis $H_1 = \{\theta = \theta_1\}$ frente a la alternativa adicional $H_2 = \{\theta \in \Theta_2\}$, $\theta_1 \notin \Theta_2$, están enlazados entre sí del modo siguiente. Supongamos que para cada θ_1 ha sido definido su conjunto $\Theta_2 = \Theta_2(\theta_1) \not\ni \theta_1$.

Teorema 1. 1) *Examinemos para cada* θ_1 *el criterio no randomizado* $\pi = \delta$ *de nivel* $1 - \varepsilon$ *para verificar la hipótesis* H_1 *frente a* H_2, *y designemos por* $\Omega(\theta_1, \varepsilon)$ *su región de aceptación de la hipótesis* H_1. *Entonces, el conjunto*

$$\Theta^*(X, \varepsilon) = \{\theta \in \Theta;\quad X \in \Omega(\theta, \varepsilon)\}$$

será un conjunto confidencial de nivel $1 - \varepsilon$.

Al contrario, si $\Theta^*(X, \varepsilon)$ *es un conjunto confidencial de nivel* $1 - \varepsilon$, *entonces el conjunto* $\Omega(\theta_1, \varepsilon) \subset \mathscr{X}^n$, *definido en* (2) *y adoptado como región de aceptación de* H_1, *determinará el criterio para verificar* $H_1 = \{\theta = \theta_1\}$ *frente a* $H_2 = \{\theta \in \Theta_2(\theta_1)\}$ *de nivel* $1 - \varepsilon$ *para cualquier* $\Theta_2(\theta_1)$, $\theta_1 \notin \Theta_2(\theta_1)$.

2) *Si el criterio* π *con la región de aceptación* $\Omega(\theta_1, \varepsilon)$ *de la hipótesis* H_1 *es el c.u.m.p., entonces, el conjunto respectivo* $\Theta^*(X, \varepsilon)$ *minimizará la probabilidad*

$$\mathbf{P}_\theta(\theta' \in \Theta^*(X, \varepsilon)) \text{ para todos } \theta, \theta', \theta \in \Theta_2(\alpha') \quad (4)$$

en la clase de todos los conjuntos confidenciales de nivel $1 - \varepsilon$.

También es cierta la afirmación contraria: La minimalidad (4) *significa que el conjunto respectivo* Ω (θ, ε) *engendrará el c.u.m.p.*

Para el parámetro unidimensional se usan principalmente los casos

$$\Theta_2(\theta') = \{\theta: \theta \neq \theta'\} \text{ y } \Theta_2(\theta') = \{\theta: \theta > \theta'\} \text{ (o bien } \{\theta: \theta < \theta'\}).$$

En el primero de ellos en (4) tendrá lugar la minimización para todos $\theta' \neq \theta$, y en el segundo, para todos $\theta' < \theta$.

Así pues, en (4), el teorema afirma que para Θ^*, la probabilidad $\mathbf{P}_\theta$ se minimiza de que todo otro valor de $\theta' \neq \theta$, tal que $\theta \in \Theta_2(\theta')$, pertenezca a un conjunto confidencial. Esta es una de las maneras de separar los intervalos confidenciales óptimos.

Definición 2. Los conjuntos confidenciales para los cuales se minimiza (4) a condición (1) se llaman *conjuntos confidenciales más exactos* (de nivel $1 - \varepsilon$) *respecto a las alternativas* θ' *tales que* $\theta \in \Theta_2(\theta')$.

Más adelante expondremos cierta argumentación adicional para tal entendimiento del intervalo confidencial óptimo.

Ahora bien, el teorema 1, establece que la "inversión" del conjunto Ω (θ_1, ε) para el c.u.m.p. da el conjunto confidencial más exacto. En este caso es importante señalar que el referido procedimiento de construcción de los conjuntos confidenciales no está de ningún modo relacionado con la dimensión de θ. Incluso se pueden examinar los parámetros de dimensión infinita θ e identificar θ con la propia distribución $\mathbf{P}$ de la muestra X. Entonces, las relaciones de equivalencia (3), donde Ω $(\theta, \varepsilon) = \Omega(\mathbf{P}, \varepsilon)$ es la región de aceptación de la hipótesis $\{X \Subset \mathbf{P}\}$ frente a la alternativa $\{X \Subset \mathbf{P}_1 \neq \mathbf{P}\}$, permiten construir el conjunto confidencial para $\mathbf{P}$. Por ejemplo, en el § 1.6 hemos visto que la distribución de la estadística $D_n = \sqrt{n} \sup_t |F_n^*(t) - F(t)|$, a condición de que $X \Subset \mathbf{P}$, donde F es una función continua de la distribución correspondiente a $\mathbf{P}$, no depende de F y puede ser determinada. Por consiguiente, podemos hallar tal $d = d(\varepsilon)$, que

$\mathbf{P}(D_n \leqslant d(\varepsilon)) = 1 - \varepsilon$. Ahora bien, la desigualdad

$$\sqrt{n} \sup_t |F_n^*(t) - F(t)| \leqslant d$$

define la región de aceptación de la hipótesis $\{X \in \mathbf{P}\}$ para el criterio de nivel $1 - \varepsilon$.

Pero esta misma desigualdad también define el conjunto confidencial para F: simplemente debido a la simetría de esta desigualdad respecto a F y F_n^*; aquí no se necesita ningún procedimiento especial de "inversión".

La **demostración del teorema** 1 es casi evidente. La misma se basa en la equivalencia (3), en virtud de la cual

$$\mathbf{P}_\theta(\theta \in \Theta^*(X, \varepsilon)) = \mathbf{P}_\theta(X \in \Omega(\theta, \varepsilon)) \geqslant 1 - \varepsilon.$$

Esto demuestra la primera afirmación. Para demostrar la segunda examinemos cualquier otro conjunto confidencial $\tilde{\Theta}^*(X, \varepsilon)$, y sea $\tilde{\Omega}(\theta, \varepsilon)$ el subconjunto correspondiente en $\mathscr{X}^n$.

Entonces,

$$\mathbf{P}_\theta(X \in \tilde{\Omega}(\theta, \varepsilon)) = \mathbf{P}_\theta(\theta \in \tilde{\Theta}^*(X, \varepsilon)) \geqslant 1 - \varepsilon,$$
$$\mathbf{P}_\theta(X \in \tilde{\Omega}(\theta_1, \varepsilon)) \geqslant \mathbf{P}_\theta(X \in \Omega(\theta_1, \varepsilon))$$

para todos $\theta \in \Theta_2(\theta_1)$ y, por lo tanto,

$$\mathbf{P}_\theta(\theta_1 \in \tilde{\Theta}^*(X, \varepsilon)) \geqslant \mathbf{P}_\theta(\theta_1 \in \Theta^*(X, \varepsilon)). \quad \triangleleft$$

Examinemos ahora un importante caso particular relacionado con el parámetro unidimensional θ.

2. Intervalos confidenciales más exactos.

Teorema 2. *Supongamos que el conjunto* $\Omega(\theta, \varepsilon)$ *del c.u.m.p. examinado en el teorema 1 tiene la forma*

$$c_1(\theta, \varepsilon) \leqslant T(x) \leqslant c_2(\theta, \varepsilon),$$

donde $c_i(\theta, \varepsilon)$ *dependen monótona y continuamente** *de* θ. Supongamos, para precisar, que $c_i(\theta, \varepsilon)$ *crecen. Entonces, el conjunto confidencial más exacto (de nivel* $1 - \varepsilon$*) respecto a las alternativas* θ' *tales, que* $\theta \in \Theta_2(\theta')$, *tendrá la forma de intervalo*

$$c_2^{-1}(T, \varepsilon) \leqslant \theta \leqslant c_1^{-1}(T, \varepsilon),$$

donde $T = T(X)$, $c_i^{-1}(t, \varepsilon)$ *son las soluciones de las ecuaciones* $c_i(\theta, \varepsilon) = t$ *respecto a* θ.

---

*) Las propiedades de monotonía y de continuidad de $c_i(\theta, \varepsilon)$ se deducen, por lo general, de las mismas propiedades de la función de distribución $\mathbf{P}_\theta T(X) < c$). En las designaciones del § 2.31, $c_1(\theta, \varepsilon) = G_\theta^{-1}(\varepsilon_1)$, $c_2(\theta, \varepsilon) = G_\theta^{-1}(1 - \varepsilon_2)$, donde G_θ es la función de distribución $T(X)$, $\varepsilon_1 + \varepsilon_2 = \varepsilon$.

Ahora bien, vemos que el procedimiento de construcción del intervalo confidencial es aquí, de hecho, el mismo que en el § 2.31 con la única particularidad de que en calidad de estadística S aquí se utiliza la estadística T del criterio uniformemente más potente.

La demostración del teorema es evidente y se la dejamos al lector.

Ahora examinemos más detalladamente los *intervalos confidenciales unilaterales para θ escalar.* Estos intervalos se utilizan allí donde reviste mayor interés una sola cota para estimar el parámetro. Tales situaciones surgen cuando se estima la probabilidad de que se produzca cualquier suceso indeseable o, digamos, cuando se estima el esfuerzo de rotura de una nueva aleación.

Debido a la simetría es posible reducirse al examen de la frontera confidencial inferior $\theta^-(X, \varepsilon)$ para la cual

$$\mathbf{P}_\theta(\theta^-(X, \varepsilon) \leqslant \theta) \geqslant 1 - \varepsilon. \tag{5}$$

Definición 3. La frontera $\theta^- = \theta^-(X, \varepsilon)$ para la cual $\mathbf{P}_\theta(\theta^- \leqslant \theta')$ es mínima con todos $\theta' < \theta$ se llama *frontera confidencial inferior más exacta de nivel* $1 - \varepsilon$.

Supongamos que $w(\theta^-, \theta)$ es cualquier medida de pérdidas que surgen debido a la "subestimación" de θ: $w(\theta^-, \theta) = 0$ cuando $\theta^- \geqslant \theta$ y $w(\theta^-, \theta) \geqslant 0$ cuando $\theta^- < \theta$; en este caso $w(\theta^-, \theta)$ crece continuamente al alejarse θ^- de θ, $\mathbf{M}_\theta w(\theta^-, \theta) < \infty$.

La siguiente afirmación aclara, en cierta medida, el sentido de la definición 3.

Lema 1. *La frontera inferior más exacta θ^- minimiza el valor $\mathbf{M}_\theta w(\theta^-, \theta)$ para la condición* (5) *y para cualquier función w que posea las propiedades enunciadas anteriormente.*

Demostración. Sea $\tilde{\theta}^-$ otra frontera inferior. Entonces, como los incrementos $d_u w(u, \theta)$ respecto a u en la región $u < \theta$ son negativos,

$$\mathbf{M}_\theta w(\theta^-, \theta) = \int_\infty^\theta w(u, \theta) d_u \mathbf{P}_\theta(\theta^- < u) = -\int_\infty^\theta \mathbf{P}_\theta(\theta^- < u) d_u w(u, \theta) \leqslant$$

$$\leqslant -\int_\infty^\theta \mathbf{P}_\theta(\tilde{\theta}^- < u) d_u w(u, \theta) = \mathbf{M}_\theta w(\tilde{\theta}^-, \theta) \quad \lhd$$

Así pues, vemos que el enfoque de la definición de los conjuntos confidenciales más exactos en caso de los conjuntos unilaterales es muy natural. Ahora, con ayuda de los teoremas 1 y 2 y los resultados del § 5 se pueden construir explícitamente los intervalos confidenciales unilaterales para el caso cuando la relación de verosimilitud es monónona.

Teorema 3. *Supongamos que $X \in \mathbf{P}_\theta$ y que la familia $\{\mathbf{P}_\theta\}$ tiene relación de verosimilitud monótona respecto a la estadística $T(X)$ cuya $\mathbf{P}_\theta$-distribución $G_\theta(t) = \mathbf{P}_\theta(T(X) < t)$ es continua respecto a θ y t. Enton-*

ces, la estadística T de la distribución depende monótona y continuamente de θ, (o sea, $G_\theta(t)$ *decrece continuamente* con el crecimiento de θ, véase la definición 2.31.3). *Si* $b(t, \gamma)$ *es la solución de la ecuación* $G_\theta(t) = \gamma$ *respecto a* θ, *entonces, la frontera inferior más exacta* $\theta^-(X, \varepsilon)$ *de nivel* $1 - \varepsilon$ *es igual a*

$$\theta^-(X, \varepsilon) = b(T(X), 1 - \varepsilon).$$

Con otras palabras, en la afirmación del teorema 2.31.1 obtendremos la frontera confidencial inferior más exacta si utilizamos en calidad de S la estadística T.

Demostración. En nuestro caso, en condiciones de los teoremas 1 y 2 es necesario poner $\Theta_2(\theta) = \{t: t > \theta\}$. En virtud del teorema 5.1 existe un c.u.m.p. no randomizado para verificar $H_1 = \{\theta = \theta_1\}$ frente a $H_2 = \{\theta > \theta_1\}$ con la región $\Omega(\theta_1, \varepsilon) = \{X: T(X) < c\}$ de aceptación de H_1, donde $c = c(\theta_1, 1 - \varepsilon) = G_{\theta_1}^{-1}(1 - \varepsilon)$ se deduce de la condición

$$\mathbf{P}_{\theta_1}(T(X) < c(\theta_1, 1 - \varepsilon)) = 1 - \varepsilon.$$

En este caso

$$\mathbf{P}_\theta(T(X)) \geqslant c) > \varepsilon = \mathbf{P}_{\theta_1}(T(X) \geqslant c)$$

cuando $\theta > \theta_1$. Esto último quiere decir que $c(\theta_1, 1 - \varepsilon) < c(\theta, 1 - \varepsilon)$ cuando $\theta_1 < \theta$, o sea, la función $c(\theta, 1 - \varepsilon)$ crece respecto a θ. La continuidad de $c(\theta, 1 - \varepsilon) = G_\theta^{-1}(1 - \varepsilon)$ respecto a θ se deduce de la continuidad de G_θ.

Vemos que las condiciones de los teoremas 1 y 2 se cumplen por completo cuando $c_2(\theta, \varepsilon) = c(\theta, 1 - \varepsilon)$ y, por lo tanto, el conjunto confidencial más exacto tiene la forma del semiintervalo $(c^{-1}(T(X), 1 - \varepsilon), \infty)$, donde, como hemos visto en el teorema 2.31.1, $c^{-1}(T, 1 - \varepsilon) = b(T, 1 - \varepsilon)$. ◁

De un modo exactamente igual se puede construir la frontera superior más exacta $\theta^+(X, \varepsilon)$.

Ahora supongamos que $\theta^-(X, \varepsilon_1) < \theta^+(X, \varepsilon_2)$ designan las fronteras confidenciales superior e inferior de los niveles $1 - \varepsilon_1$ y $1 - \varepsilon_2$, respectivamente. Como los sucesos $\{\theta^-(X, \varepsilon_1) > \theta\}$ y $\{\theta^+(X, \varepsilon_2) < \theta\}$ son disjuntos, entonces

$$\mathbf{P}_\theta(\theta^-(X, \varepsilon_1) < \theta < \theta^+(X, \varepsilon_2)) = 1 - \varepsilon_1 - \varepsilon_2,$$

y $(\theta^-(X, \varepsilon_1), \theta^+(X, \varepsilon_2)$ es el intervalo confidencial de nivel $1 - \varepsilon_1 - \varepsilon_2$.

Sean $w_1(\theta^-, \theta)$ y $w_2(\theta^+, \theta)$ las funciones de pérdidas para las fronteras $\theta^\mp$ que poseen las propiedades descritas en la enunciación del lema 1.

Lema 2. *Sea* $w(\theta^-, \theta^+, \theta) = w_1(\theta^-, \theta) + w(\theta^+, \theta)$. *Entonces, el intervalo confidencial* (θ^-, θ^+), *formado por las fronteras superiores e inferiores más exactas, minimiza* $\mathbf{M}_\theta w(\theta^-, \theta^+, \theta)$ *para las condiciones*

$$\mathbf{P}_\theta(\theta^- > \theta) \leqslant \varepsilon_1, \ \mathbf{P}_\theta(\theta^+ < \theta) \leqslant \varepsilon_2.$$

Este lema es el corolario evidente del lema 1. El mismo muestra que el intervalo confidencial construido con ayuda de las fronteras inferior exacta y superior exacta también poseerá propiedades de optimización.

El teorema 3 da la posibilidad de construir explícitamente tales intervalos para las familias paramétricas que tienen monótonas las relaciones de verosimilitud.

Le proponemos al lector que el mismo se cerciore, a base de las observaciones efectuadas, de que los intervalos confidenciales, construidos en el § 2.32 para la media y la varianza de la distribución normal, tendrán las fronteras superiores e inferiores más exactas.

En el teorema 1 y en las investigaciones posteriores figuraba la condición de que el c.u.m.p. no es randomizado. Sin embargo, esta limitación no es importante. Cualquier criterio randomizado π puede ser representado como criterio no randomizado, si en la investigación se introduce una observación adicional Y que sea independiente de X y que esté uniformemente distribuida en [0, 1]. En efecto, examinemos, para la nueva muestra (X, Y), la región crítica

$$\Omega = \{(x, y): \pi(x) \geqslant y\},$$

o sea, supongamos que $\delta(X, Y) = 1$ si $(X, Y) \in \Omega$, y que $\delta(X, Y) = 0$ en el caso contrario. Entonces, para toda distribución de X,

$$\mathbf{P}(\delta(X, Y) = 1) = \mathbf{P}(\pi(X) \geqslant Y) = \int_0^1 \mathbf{P}(\pi(X) \geqslant y)dy = \mathbf{M}\pi(x),$$

y, por consiguiente, el criterio δ es equivalente (según sus parámetros) a π. ¿Cómo aprovechar esta circunstancia para construir los intervalos confidenciales en condiciones del teorema 3? Supongamos, para abreviar, que la estadística $T(X)$ es de números enteros (como hemos visto, la falta de los c.u.m.p. sólo puede ser provocada por el carácter discreto de la distribución T). Entonces, la observación $S(X, Y) = T(X) + Y$, $Y \in U_{0,1}$ conserva toda la información contenida en $T(X)$, ya que $T(X)$ es una parte entera de $S(X, Y)$. Eligiendo $c(\theta, \varepsilon)$ entero, al c.u.m.p. de nivel $1 - \varepsilon$ se le puede conferir la forma siguiente: se acepta la hipótesis H_1 si

$$S(X, Y) \leqslant c(\theta_1, 1 - \varepsilon).$$

Así pues, hemos construido los conjuntos requeridos $\Omega(\theta, \varepsilon)$ y sólo queda "invertirlos" usando el mismo procedimiento que antes. Obtendremos la frontera inferior

$$\theta^-(X, Y, \varepsilon) = c^{-1}(T(X) + Y, 1 - \varepsilon),$$

donde c^{-1} es la función inversa a c con arreglo al primer argumento. Aquí, de la propia escritura se deduce que para definir θ^- es necesario realizar una observación adicional Y.

Ejemplo 1. Sea $X \in \mathbf{B}_p$, y nos interesa la frontera confidencial superior p^+ de nivel $1 - \varepsilon$ para la probabilidad $p = \mathbf{P}(x_i = 1) = 1 - \mathbf{P}(x_i = 0)$. La familia de distribuciones $\{\mathbf{B}_p\}$ es exponencial y satisface las condiciones del teorema 3, donde conviene poner $T(X) = \sum_{i=1}^{n} x_i$. Examinemos la observación

$$S = \sum_{i=1}^{n} x_i + Y, \quad Y \in U_{0,1}.$$

Esta tiene en el punto t, $0 \leqslant t \leqslant n + 1$, la densidad $C_n^{[t]} p^{[t]}(1 - p)^{n-[t]}$. Designemos por $Gp^{(t)}$ la función de distribución con esta densidad. Entonces p^+ será la solución de la ecuación $G_p(t) = \varepsilon$.

3. Conjuntos confidenciales no desplazados. Volvamos a la cuestión acerca de los conjuntos confidenciales más exactos. Con ayuda del teorema 3 podemos construir las fronteras superiores e inferiores más exactas basándose en el hecho de que para las alternativas unilaterales $\{\theta > \theta_1\}$, $\{\theta < \theta_1\}$ de las hipótesis $\{\theta = \theta_1\}$, en una serie de casos existe el c.u.m.p. Si tratamos de utilizar los teoremas 1 y 2 directamente para construir los intervalos confidenciales más exactos, necesitaremos la existencia de c.u.m.p. para verificar la hipótesis $\{\theta = \theta_1\}$ frente a $\{\theta \neq \theta_1\}$, lo cual ocurre múy raramente. La salida de esta posición consiste en la reducción natural de la clase de intervalos confidenciales sujetos a investigación, procediendo del mismo modo que cuando reducimos las clases de criterios examinados en el § 6.7, es decir, introduciendo los conceptos de conjuntos confidenciales no desplazados e invariantes.

Supongamos que, como antes, a cada θ le corresponde el conjunto $\Theta_2(\theta)$, $\theta \notin \Theta_2(\theta)$.

Definición 4. El conjunto confidencial $\Theta^*(X, \varepsilon)$ para θ de nivel $1 - \varepsilon$ se considera *no desplazado respecto a las alternativas* θ', *tales que* $\theta \in \Theta_2(\theta')$ si

$$\mathbf{P}_\theta(\theta' \in \Theta^*(X, \varepsilon)) \leqslant 1 - \varepsilon \text{ para todos } \theta,\ \theta',\ \theta \in \Theta_2(\theta'). \qquad (6)$$

El conjunto $\Theta^*(X, \varepsilon)$ se considera simplemente *no desplazado* si (6) es válida para todos $\theta' \neq \theta$.

El no desplazamiento del conjunto confidencial significa que la *probabilidad de que éste recubra el valor falso de* θ' *no es mayor que la probabilidad de que el mismo recubra el valor verdadero.*

Definición 5. Los conjuntos confidenciales para los cuales se minimiza (4) en condiciones (1) y (6) se llaman *conjuntos confidenciales no desplazados más exactos* ((de nivel $1 - \varepsilon$) respecto a las alternativas para las cuales $\theta \in \Theta_2(\theta')$.

Teorema 4. 1) *Los criterios no randomizados y no desplazados engendran, en virtud de la equivalencia (3), conjuntos confidenciales no desplazados, y al contrario.*

2) *Si* $\Omega(\theta_1, \varepsilon)$ *para cada* $\theta_1 \in \Theta$ *es la región de aceptación de la hipótesis* $\{\theta = \theta_1\}$ *del criterio uniformemente más potente no desplazado y no randomizado, con una alternativa* $\{\theta \in \Theta_2(\theta_1)\}$, *entonces, el conjunto respectivo* $\Theta^*(X, \varepsilon)$ *será el conjunto confidencial no desplazado más exacto, y al contrario.*

La demostración del teorema repite por completo los razonamientos del teorema 1, a los cuales sólo es necesario añadir que la propiedad de desplazamiento se conserva al pasar de los criterios a los conjuntos confidenciales y al contrario. En efecto, las relaciones (1) y (6) son equivalentes a

$$\sup_{\theta \in \Theta_2(\theta_1)} \mathbf{P}_\theta(X \in \Omega(\theta_1, \varepsilon)) \leqslant 1 - \varepsilon \leqslant \mathbf{P}_{\theta_1}(X \in \Omega(\theta_1, \varepsilon)).$$

Si $\pi(X)$ es la función crítica de los criterios no randomizados que figuran en el teorema $\pi(X) = 0$ para $X \in \Omega(\theta_1, \varepsilon))$, entonces obtenemos

$$\mathbf{M}_\theta \pi(X) = 1 - \mathbf{P}_\theta(X \in \Omega(\theta_1, \varepsilon)),$$
$$\inf_{\theta \in \Theta_2(\theta_1)} \mathbf{M}_\theta \pi(X) \geqslant \varepsilon \geqslant \mathbf{M}_{\theta_1} \pi(X).$$

Esta es, precisamente, la propiedad de no desplazamiento que equivale a (6). ◁

Si utilizamos los resultados del § 6 y construimos el conjunto confidencial no desplazado y más exacto para el parámetro θ de una familia exponencial, obtendremos el mismo intervalo confidencial (θ^-, θ^+) que hemos construido utilizando la monotonía de la relación de verosimilitud, o sea, el intervalo en el cual θ^- y θ^+ son las fronteras inferior y superior más exactas, respectivamente, de niveles $1 - \varepsilon/2$.

4. Conjuntos confidenciales invariantes. La siguiente definición utiliza las designaciones y los conceptos del párrafo precedente. Sea $\{\mathbf{P}_\theta\}$ una familia invariante respecto a G.

Definición 6. El conjunto confidencial $\Theta^*(X, \varepsilon)$ se llama *invariante**) respecto al grupo G si

$$\Theta^*(gX, \varepsilon) = \bar{g}\Theta^*(X, \varepsilon) \tag{7}$$

para todos $g \in G$.

El sentido de este concepto es análogo al de la estimación equivariante (§ 2.19). Si las transformaciones g y $\bar{g}$ se interpretan como la sustitución del sistema de coordenadas que conserva la distribución, entonces (7) significará que el conjunto confidencial no depende del sistema de coordenadas en el que se expresan los datos iniciales.

---

*) Ateniéndose a la observación expuesta en la p. 195 del § 2.19, sería más natural llamar el conjunto confidencial con propiedad (7), conjunto equivariante.

Definición 7. El conjunto confidencial $\Theta^*(X, \varepsilon)$ *se denomina conjunto confidencial invariante más exacto* de nivel $1 - \varepsilon$, si en él se minimiza $\mathbf{P}_\theta(\theta' \in \Theta^* (X, \varepsilon))$ para todos $\theta' \neq \theta$ en la clase de todos los conjuntos Θ^* que satisfacen (7) y la condición $\mathbf{P}_\theta(\theta \in \Theta^*(X, \varepsilon) = 1 - \varepsilon$. Sea Ω (θ_1, ε) la región de aceptación de la hipótesis $H_1 = \{\theta = \theta_1\}$ cuando la alternativa constituye $\{\theta \neq \theta_1\}$ para el criterio invariante de nivel $1 - \varepsilon$. Nótese que hay una diferencia esencial en las definiciones del criterio invariante y del conjunto confidencial invariante (esta diferencia no existiría si se necesitara el cumplimiento de la igualdad $g\Omega(\theta, \varepsilon) = \Omega(\bar{g}\theta, \varepsilon)$ y no de la igualdad $g\Omega(\theta, \varepsilon) = \Omega$ (θ, ε)). Con este hecho está relacionada la circunstancia de que la correspondencia entre los criterios invariantes uniformemente más potentes y los intervalos confidenciales invariantes más exactos tiene un aspecto más complejo que en los teoremas precedentes.

Examinemos el grupo de transformaciones G y supongamos que para cada θ en este grupo hay un subgrupo $G[\theta_1]$ que deja invariante el problema de verificación de la hipótesis $H_1 = \{\theta = \theta_1\}$. Con otras palabras, $\bar{g}\theta_1 = \theta_1$ cuando $g \in G[\theta_1]$.

Teorema 5. *Sea* $\Theta^*(X, \varepsilon)$ *un conjunto confidencial de nivel* $1 - \varepsilon$ *invariante respecto a* G. *Entonces*

1) *La región* Ω $(\theta, \varepsilon) = \{x: \theta \in \Theta^* (x, \varepsilon)\}$ *será invariante respecto a* $G[\theta]$ *para cada* θ.

2) *Si la región* Ω (θ_1, ε), correspondiente a $\Theta^*(X, \varepsilon)$, es *la región de aceptación de* H_1 *cuando la alternativa constituye* $\{\theta \neq \theta_1\}$ *para el criterio invariante uniformemente más potente de nivel* $1 - \varepsilon$, *entonces* $\Theta^*(X, \varepsilon)$ *será el conjunto invariante confidencial más exacto.*

Demostración. 1) Supongamos que $g \in G[\theta]$. Entonces $\bar{g}\theta = \theta$,

$$g\Omega(\theta, \varepsilon) = \{gx: \theta \in \Theta^*(x, \varepsilon)\} = \{x: \theta \in \Theta^*(g^{-1}x, \varepsilon)\} =$$
$$= \{x: \theta \in \bar{g}^{-1}\Theta^*(x, \varepsilon)\} = \{x: \bar{g}\theta \in \Theta^*(x, \varepsilon)\} =$$
$$= \{x: \theta \in \Theta^*(x, \varepsilon)\} = \Omega(\theta, \varepsilon).$$

2) Sea $\tilde{\Theta}^*$ cualquier otro conjunto confidencial invariante de nivel $1 - \varepsilon$. Según la primera afirmación, a él le corresponde el criterio invariante de nivel $1 - \varepsilon$ con la región $\tilde{\Omega}$ (θ_1, ε) de aceptación de H_1.

Como, por suposición,

$$\mathbf{P}_\theta(X \in \Omega\ (\theta_1, \varepsilon)) \geqslant \mathbf{P}_\theta(X \in \tilde{\Omega}(\theta_1, \varepsilon)),$$

entonces

$$\mathbf{P}_\theta(\theta_1 \in \Theta^*(X, \varepsilon)) \geqslant \mathbf{P}_\theta(\theta_1 \in \tilde{\Theta}^*(X, \varepsilon)).$$

cuando $\theta_1 \neq \theta$. Que es lo que se necesitaba demostrar. ◁

Ejemplo 2. Supongamos que $X \in \Phi_{\alpha,\sigma^2}$. Se necesita construir el conjunto confidencial más exacto para el parámetro σ^2, siendo desconocido α. En el ejemplo 2 del párrafo precedente hemos visto que la familia Φ_{α,σ^2} es invariante respecto a las transformaciones de desplazamiento $gX = X + c$ si $\bar{g}(\alpha, \sigma^2) = (\alpha + c, \sigma^2)$. La estadística $S_0^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ es el máximo invariante construido según la estadística suficiente. Además, la hipótesis $H_1 = \{\sigma = \sigma_1\}$ es invariante respecto a G. Conforme al ejemplo 7.2, el criterio uniformemente más potente invariante y no desplazado para verificar H_1 tiene la forma

$$h_{1,\varepsilon}\sigma_1^2 < (n-1)S_0^2 < h_{2,\varepsilon}\sigma_1^2, \tag{8}$$

donde $h_{i,\varepsilon}$ se deduce de las condiciones (véase la condición (6.7) del teorema 6.1):

$$\mathbf{P}(h_{1,\varepsilon} < \chi^2_{n-1} < h_{2,\varepsilon}) = 1 - \varepsilon,$$
$$\mathbf{M}(\chi^2_{n-1}; h_{1,\varepsilon} < \chi^2_{n}-1 < h_{2,\varepsilon}) = (1-\varepsilon)\mathbf{M}\chi^2_{n-1},$$
$$\chi^2_{n-1} \in H_{n-1}.$$

El conjunto confidencial $\Theta^*(X, \varepsilon)$ correspondiente a (8) tiene la forma del intervalo

$$(n-1)S_0^2/h_{2,\varepsilon} < \sigma^2 < (n-1)S_0^2/h_{1,\varepsilon}. \tag{9}$$

Este intervalo es, evidentemente, invariante respecto a g, al igual que el criterio (8) (en este ejmplo $G|\sigma_1| \equiv G$ para cualquier σ_1). Por lo tanto, en virtud de las segundas afirmaciones de los teoremas 4 y 5, el intervalo (9) es el conjunto confidencial no desplazado e invariante más exacto de nivel $1 - \varepsilon$.

Ejemplo 3. Supongamos que $X \in \Phi_{\alpha\sigma^2}$. Es necesario construir el conjunto confidencial máx exacto para el parámetro α cuando se desconoce σ. Aquí

$$f_{\alpha,\sigma^2}(X) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \alpha)^2\right\}.$$

La familia Φ_{α,σ^2} será invariante respecto al grupo G de las transformaciones lineales $gX = aX + b$ si se pone $\bar{g}(\alpha, \sigma) = (a\alpha + b, =a\sigma)$. El par de observaciones $(\bar{x}, S_0^2)$ forma una estadística suficiente. Es fácil ver que con su ayuda no se puede construir una estadística que sea invariante respecto a G. No obstante, para cada α_1 se puede separar un subgrupo $G[\alpha_1]$ de transformaciones $gX = a(X - \alpha_1) + \alpha_1$ respecto al cual la estadística $(\bar{x} - \alpha_1)/S_0$ será el máximo invariante. La hipótesis $H_1 = \{\alpha = \alpha_1\}$ queda invariante respecto a $G[\alpha_1]$. Investigando la densidad $(\bar{x} - \alpha_1)/S_0$ se puede

mostrar, con ayuda de los métodos del § 7 (omitimos estas consideraciones puesto que son muy complicadas*), que para cada σ, el criterio uniformemente más potente no desplazado e invariante para verificar la hipótesis H_1 frente a $\{\alpha \neq \alpha_1\}$ existe y tiene una región de aceptación de H_1 en forma de

$$\sqrt{n}|\bar{x} - \alpha_1|/S_0 < \tau_\varepsilon, \tag{10}$$

donde τ_ε se determina de la condición $\mathbf{P}(|t_{n-1}| \geqslant \tau_\varepsilon) = \varepsilon$, $t_{n-1} \in T_{n-1}$.

El conjunto confidencial respectivo Θ^* tiene la forma

$$\bar{x} - \tau_\varepsilon S_0/\sqrt{n} < \alpha < \bar{x} + \tau_\varepsilon S_0/\sqrt{n}. \tag{11}$$

Es fácil ver que este intervalo confidencial es invariante ($\Theta^*(gX, \varepsilon) = \bar{g}\Theta^*(X, \varepsilon)$). Según la primera afirmación del teorema 5, el criterio (10) será invariante respecto a $G[\alpha_1]$. De acuerdo con la segunda afirmación, el intervalo confidencial (11) será el criterio confidencial más exacto (uniformemente respecto a σ) no desplazado e invariante de nivel $1 - \varepsilon$.

Ahora bien, en este párrafo hemos establecido que todos los intervalos confidenciales construidos en el § 2.32 son, en cierto sentido, óptimos.

§ 9. Enfoques bayesiano y minimax de la verificación de las hipótesis compuestas

1. Criterios bayesianos y minimax. En el § 4 hemos descrito los enfoques bayesiano y minimax. Allí mismo hemos dado las definiciones respectivas que recordaremos en la exposición posterior.

Supongamos, como antes, que se verifica la hipótesis $H_1 = \{\theta \in \Theta_1\}$ frente a $H_2 = \{\theta \in \Theta_2\}$, basándose en la muestra $X \in \mathbf{P}_\theta$.

El *enfoque bayesiano completo* supone que Θ se elige al azar con la distribución a priori $\mathbf{Q}$ en $\Theta = \Theta_1 \cup \Theta_2$. La distribución $\mathbf{Q}$ induce las distribuciones $\mathbf{Q}_i$ en Θ_i, $i = 1, 2$ y las probabilidades $q(i) = \mathbf{Q}(\theta \in \Theta_i)$, así que $\mathbf{Q} = q(1)\mathbf{Q}_1 + q(2)\mathbf{Q}_2$. Designemos por H_{Q_i} la hipótesis de que $\theta \in \Theta_i$ se elige al azar, con la distribución $\mathbf{Q}_i$. Según esta hipótesis, X tiene la densidad

$$f_{Q_i}(x) = \int f_\theta(x)\mathbf{Q}_i(dx).$$

Se entiende, por supuesto (véase el § 4), que en Θ_i están definidas las σ-álgebras de σ_i, a base de las cuales se eligen $\mathbf{Q}_i$, y que $f_\theta(x)$ es medible respecto a $\mathfrak{S}_i \times \mathfrak{B}^n_{\mathscr{X}}$.

De los resultados del § 1, 2 se deduce que el criterio bayesiano π_Q para verificar H_{Q_1} frente a H_{Q_2} en el problema descrito anteriormente tendrá

*) Esto se expone más detalladamente en [57], p. 312.

la forma

$$\pi_Q(X) = \begin{cases} 1, & \text{si } f_{Q_2}(X) > cf_{Q_1}(X), \\ p, & \text{si } f_{Q_2}(X) = cf_{Q_1}(X), \\ 0, & \text{si } f_{Q_2}(X) < cf_{Q_1}(X), \end{cases} \tag{1}$$

donde $c = q(1)/q(2)$, $p \in [0, 1]$ es arbitrario.

El *enfoque parcialmente bayesiano* está relacionado con la verificación de la hipótesis H_{Q_1} frente a H_{Q_2} en el caso cuando falta la distribución a priori entre H_{Q_1} y H_{Q_2} (que se define por las probabilidades $q(1)$ y $q(2)$). Pongamos

$$K_\varepsilon^{Q_1} = \{\pi: \mathbf{M}_{Q_1}\pi(X) \leqslant \varepsilon\}.$$

Entonces el criterio $\pi_{Q_1Q_2}$ se llama bayesiano en $K_\varepsilon^{Q_1}$ si éste es el c.m.p. de nivel $1 - \varepsilon$ para verificar H_{Q_1} frente a H_{Q_2}. El criterio $\pi_{Q_1Q_2}$ tendrá la misma forma (1), donde c y p se eligen de la condición $\mathbf{M}_{Q_1}\pi_{Q_1Q_2}(X) = \varepsilon$.

En vez de $\pi_{Q_1Q_2}$ escribiremos π_{Q_1} y π_{Q_2} si uno de los conjuntos Θ_1 o Θ_2 se degenera en conjunto de un punto $\{\theta_1\}$ o $\{\theta_2\}$.

En las aplicaciones rara vez se encuentran problemas en las que las distribuciones $\mathbf{Q}_i$ son completamente conocidas. Sin embargo, ya hemos visto repetidas veces que la utilidad del enfoque bayesiano no se limita exclusivamente a la posibilidad de aplicarlo directamente. Este enfoque permite construir los c.u.m.p., y también los minimax (compárese con los §§ 1, 5 y 6). Posteriormente utilizaremos el enfoque bayesiano también para construir los criterios asintóticamente óptimos. Sea, como antes,

$$K_\varepsilon = \{\pi: \sup_{\theta\in\Theta_1} \mathbf{M}_\theta\pi(X) \leqslant \varepsilon\}. \tag{2}$$

Entonces el criterio, $\bar{\pi}$ se denomina minimax en K_ε (en $K_\varepsilon^{Q_1}$) si $\bar{\pi} \in K_\varepsilon$ ($\bar{\pi} \in K_\varepsilon^{Q_1}$), y para él se minimiza

$$\inf_{\theta\in\Theta_2} \mathbf{M}_\theta\pi(X) = \inf_{\theta\in\Theta_2} \beta(\theta). \tag{3}$$

Cabe señalar que si las funciones de potencia $\beta(\theta) = \mathbf{M}_\theta\pi(X)$ son continuas y los conjuntos Θ_1 y Θ_2 se tocan, entonces

$$\beta = \sup_{\pi\in K_\varepsilon} \inf_{\theta\in\Theta_2} \beta(\theta) \leqslant \varepsilon \tag{4}$$

y la desigualdad $\beta > \varepsilon$ no puede cumplirse. Por eso, si se desea que la potencia garantizada (3) sea suficientemente grande (en todo caso, mayor que ε), conviene examinar los conjuntos "separados" Θ_1 y Θ_2. Con otras palabras, es necesario eliminar la zona de los valores de θ, donde $\beta(\theta)$ es próxima a ε como zona de "indiferencia" de los criterios, y examinar, en calidad de Θ_2, el conjunto que no toca Θ_1.

23—8030

No obstante, si los conjuntos se tocan, *todo criterio no desplazado en* K_ε *será minimax*. En efecto, para los criterios no desplazados $\beta(\theta) = \mathbf{M}_\theta \pi(X) \geqslant \varepsilon$, $\theta \in \Theta_2$ y, por lo tanto, $\beta = \inf_{\theta \in \Theta_2} \beta(\theta) \geqslant \varepsilon$ alcanza, en virtud de (4), su valor máximo.

La afirmación inversa es cierta en el caso general: *el criterio minimax, si existe, no está desplazado*. Esto se desprende del hecho de que

$$\beta = \sup_{\pi \in K_\varepsilon} \inf_{\theta \in \Theta_2} \beta(\theta) \geqslant \varepsilon$$

(podemos tomar $\pi(X) \equiv \varepsilon$) y del hecho de que para el criterio minimax

$$\inf_{\theta \in \Theta_2} \beta(\theta) = \beta.$$

El criterio uniformemente más potente no desplazado $\breve{\pi}$ *en la clase* $\breve{K}_\varepsilon$ *de todos los criterios no desplazados, es minimax en* K_ε. En efecto, sea $\breve{\beta}(\theta)$ la función de potencia del criterio $\breve{\pi}$. Entonces, para cualesquiera $\pi \in \breve{K}_\varepsilon$, $\theta \in \Theta_2$,

$$\breve{\beta}(\theta) \geqslant \beta(\theta), \quad \inf_{\theta \in \Theta_2} \beta(\theta) \geqslant \inf_{\theta \in \Theta_2} \beta(\theta),$$

$$\inf_{\theta \in \Theta_2} \breve{\beta}(\theta) = \sup_{\pi \in \breve{K}_\varepsilon} \inf_{\theta \in \Theta_2} \beta(\theta) = \sup_{\pi \in K_\varepsilon} \inf_{\theta \in \Theta_2} \beta(\theta). \tag{5}$$

La última igualdad se explica por el hecho de que la adición a $\breve{K}_\varepsilon$ de los criterios de K_ε, para los cuales $\inf_{\theta \in \Theta_2} \beta(\theta) < \varepsilon$, no cambia la magnitud $\sup_{\pi \in \breve{K}_\varepsilon}$ en (5). ◁

En el teorema 5.3 hemos utilizado los criterios bayesianos para determinar el c.u.m.p. La siguiente afirmación es cierto "desarrollo" del teorema 5.3. La misma también es el análogo de los teoremas 1.2 y 2.11.2 y establece que los criterios minimax han de buscarse en la clase de criterios (1) cuya forma explícita conocemos.

Teorema 1. *Supongamos que existen las distribuciones* $\mathbf{Q}_i$ *concentradas, respectivamente, en los conjuntos* $\Theta_i^\circ \subset \Theta_i$, $i = 1, 2$, *y las constantes* c y p *tales, que el criterio* $\pi_{Q_1Q_2}$, *definido en* (1), *poseen las propiedades*

$$1)\ \pi_{Q_1Q_2} \in K_\varepsilon^{Q_1},$$

$$2)\ \mathbf{M}_\theta \pi_{Q_1Q_2}(X) = \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi_{Q_1Q_2}(X) \tag{6}$$

para todos $\theta \in \Theta_1^\circ$,

$$3)\ \mathbf{M}_\theta \pi_{Q_1Q_2}(X) = \inf_{\theta \in \theta_2} \mathbf{M}_\theta \pi_{Q_1Q_2}(X) \tag{7}$$

para todos $\theta \in \Theta_2^\circ$.

Entonces $\pi_{Q_1Q_2} \in K_\varepsilon$ *es precisamente el criterio minimax en* K_ε *para verificar* H_1 *frente a* H_2.

El par de distribuciones $\mathbf{Q}_1$ y $\mathbf{Q}_2$ *que posee las propiedades* 2) y 3) *es el menos favorable en el sentido de que para cualesquiera dos otras distribuciones* $\mathbf{Q}_1'$ y $\mathbf{Q}_2'$,

$$\inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi_{Q_1Q_2} \leqslant \inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi_{Q_1'Q_2'},$$

donde $\pi_{Q_1'Q_2'}$ *es el criterio de forma* (1) *de* K_ε.

La última afirmación significa que entre todos los criterios bayesianos (1), el criterio $\pi_{Q_1Q_2}$ posee la potencia menos garantizada.

Demostración. Como

$$\sup_{\theta\in\Theta_1} \mathbf{M}_\theta \pi_{Q_1Q_2}(X) = \int_{\Theta_1} \mathbf{M}_\theta \pi_{Q_1Q_2} Q_1(d\theta) = \mathbf{M}_{Q_1}\pi_{Q_1Q_2} = \varepsilon,$$

entonces $\pi_{Q_1Q_2} \in K_\varepsilon$. La potencia garantizada $\pi_{Q_1Q_2}$ es igual a (véase (7))

$$\inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi_{Q_1Q_2}(X) = \int_{\Theta_2} \mathbf{M}_\theta \pi_{Q_1Q_2} Q_2(d\theta) = \mathbf{M}_{Q_2}\pi_{Q_1Q_2} = \beta_{Q_1Q_2}. \quad (8)$$

Sea ahora π cualquier otro criterio de K_ε para verificar H_1 frente a H_2. Entonces π será simultáneamente el criterio de $K_\varepsilon^{Q_1}$ para verificar H_{Q_1} frente a H_{Q_2}, ya que

$$\mathbf{M}_{Q_1}\pi(X) = \int_{\Theta_1} \mathbf{M}_\theta \pi(X)\, Q_1(d\theta) \leqslant \sup_{\theta\in\Theta_1} \mathbf{M}_\theta \pi(X) \leqslant \varepsilon. \quad (9)$$

Pero el criterio $\pi_{Q_1Q_2}$ es el c.m.p. en $K_\varepsilon^{Q_1}$ para verificar H_{Q_1} frente a H_{Q_2}. Por consiguiente, en virtud de (8),

$$\inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi_{Q_1Q_2}(X) = \beta_{Q_1Q_2} \geqslant \mathbf{M}_{\theta_2}\pi(X) \geqslant \inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi(X). \quad (10)$$

La primera afirmación del teorema queda demostrada. Sean ahora $\mathbf{Q}_1'$ y $\mathbf{Q}_2'$ cualesquiera dos otras distribuciones en Θ_1 y Θ_2, respectivamente. El criterio $\pi_{Q_1'Q_2'}$, al igual que $\pi_{Q_1Q_2}$, será el criterio de $K_\varepsilon^{Q_1'}$ para verificar $H_{Q_1'}$ frente a $H_{Q_2'}$, ya que

$$\mathbf{M}_{Q_1'}\pi_{Q_1Q_2}(X) = \int_{\Theta_1} \mathbf{M}_\theta \pi_{Q_1Q_2}(X) Q_1'(d\theta) \leqslant \sup_{\theta\in\Theta_1} \mathbf{M}_\theta \pi_{Q_1Q_2}(X) \leqslant \varepsilon.$$

Pero el criterio $\pi_{Q_1'Q_2'}$ es el c.m.p. para estas hipótesis, por eso, en virtud de (8),

$$\beta_{Q_1'Q_2'} = \mathbf{M}_{Q_2'}\pi_{Q_1'Q_2'}(X) \geqslant \mathbf{M}_{Q_2'}\pi_{Q_1Q_2}(X) =$$
$$= \int_{\Theta_2} \mathbf{M}_\theta \pi_{Q_1Q_2}(X) Q_2'(d\theta) \geqslant \inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi_{Q_1Q_2}(X) = \beta_{Q_1Q_2}. \quad \triangleleft$$

La principal dificultad en la aplicación del teorema 1 a los problemas reales consiste en buscar (o adivinar) las distribuciones menos favorables $\mathbf{Q}_1$ y $\mathbf{Q}_2$. En este caso a veces pueden resultar útiles las consideraciones de invariación, así como ocurre en los ejemplos del apartado siguiente. Estos ejemplos tienen interés autónomo y se utilizarán posteriormente.

2. Criterios minimax para el parámetro α de distribuciones normales.

Ejemplo 1. Supongamos que $X = x_1 \in \Phi_{\alpha,E}$ es una muestra de volumen $n = 1$ de una distribución normal m-dimensional con media $\alpha = (\alpha_1, \ldots, \alpha_m)$ y con matriz unidad de segundos momentos. Designemos $|\alpha^2| = \sum_{i=1}^{m} \alpha_i^2$ y examinemos el problema de verificación de la hipótesis $H_1 = \{|\alpha| \leqslant a\}$ frente a $H_2 = \{|\alpha| \geqslant b\}$, $b > a$ (aquí hay una zona "separadora" $a < |\alpha| < b$).

Si, por ejemplo X determina (en un canal de comunicación) las amplitudes de la señal vectorial compuesta por el "ruido" $X_0 \in \Phi_{0,1}$ y por la señal útil α, $|\alpha| \geqslant b$, las hipótesis H_i se pueden considerar, para $a = 0$, como hipótesis de la presencia de la señal útil.

En vista de que el ejemplo sujeto a examen se utilizará repetidas veces posteriormente, la afirmación referente a la forma del criterio minimax será enunciada en forma de teorema.

Teorema 2. *El criterio minimax $\bar{\pi} \in K_\varepsilon$ para verificar $H_1 = \{|\alpha|\} \leqslant a\}$ frente a $H_2 = \{|\alpha|\} \geqslant b\}$, $a < b$, según la observación $X \in \Phi_{\alpha,E}$,* tiene la forma

$$\pi(X) = \begin{cases} 1, & si\ |X| > c_\varepsilon, \\ 0, & si\ |X| \leqslant c_\varepsilon, \end{cases}$$

donde c_ε se elige de la condición $p_c(a) = \varepsilon$, la potencia garantizada $\bar{\pi}$ es igual a $p_{c_\varepsilon}(b)$,

$$p_c(t) = \mathbf{P}((\xi_1 - t)^2 + \xi_2^2 + \ldots + \xi_m^2 > c^2),$$

$\xi_i \in \Phi_{0,1}$ son independientes.

Demostración. Comencemos por consideraciones sugestivas. En nuestro caso, para $x = (x^{(1)}, \ldots, x^{(m)})$ tenemos

$$f_\alpha(x) = \frac{1}{(2\pi)^{m/2}} \exp\left\{-\frac{1}{2}(x - \alpha)(x - \alpha)^T\right\},$$

donde x^T es el vector columna. De aquí se deduce que la familia de distribuciones espuesta a examen es invariante respecto a la transformación ortogonal $gx = xC$, donde C es la matriz de la transformación ortogonal en R^m. En este caso hay que poner $\bar{g}\alpha = \alpha C$. Las hipótesis H_i serán invariantes respecto a $\bar{g}$.

Supongamos, para abreviar, que $a = 0$. Si la distribución $\mathbf{Q}_2$ en $\Theta_2 = \{\alpha: |\alpha| \geqslant b\}$ no manifestara invariación respecto a $\bar{g}$ (así sucederá, por ejemplo, cuando la misma se halle concentrada en el entorno de cualquier punto α_0), entonces, esta asimetría podría utilizarse, de una u otra manera, para resolver tal problema (con la suposición que acabamos de hacer estaríamos próximos al problema de verificación de dos hipótesis simples

$\{\alpha = 0\}$ y $\{\alpha = \alpha_0\}$ y en este caso obtendríamos un criterio de gran potencia). Por lo tanto, dicha distribución no puede ser la menos favorable. Esta debe ser la distribución Q_2, invariante respecto a $\bar{g}$. Además, está claro que obtendremos la peor variante si toda la distribución permanece concentrada en la frontera Θ_2 (cuanto más semejantes sean las hipótesis, tanto más difícil será distinguirlas). Se pueden citar razonamientos sugestivos análogos respecto a Q_1, si $a \neq 0$.

Así pues, es natural que en nuestro ejemplo las distribuciones menos favorables Q_1 y Q_2 sean distribuciones uniformes en las esferas $\Theta_1 = \{\alpha: |\alpha| = a\}$ y $\Theta_2^\circ = \{\alpha: |\alpha| = b\}$. En este caso, de acuerdo con el teorema 1 el criterio minimax $\bar{\pi}$ tendrá la forma $\bar{\pi}(x) = \pi_{Q_1Q_2}(x)$, donde $\pi_{Q_1Q_2}(x) = 1$ si

$$\int_{\Theta_2^\circ} \exp\left\{-\frac{1}{2}(x-v)(x-v)^T\right\}\frac{dV(v)}{V_2} > c\int_{\Theta_1^\circ} \exp\left\{-\frac{1}{2}(x-v)(x-v)^T\right\}\frac{dV(v)}{V_1} \tag{11}$$

y $\pi_{Q_1Q_2}(x) = 0$ en el caso contrario. Aquí $dV(v)$ significa el área del elemento de la esfera correspondiente, $V_i = \text{mes}\,\Theta_i^\circ$, $i = 1, 2$.

Examinemos cualquiera de estas integrales, por ejemplo, la derecha, y notemos que ésta puede ser escrita en la forma

$$\exp\left\{-\frac{1}{2}xx^T - a^2\right\}\cdot\int_{\Theta_1^\circ}\exp\{xv^T\}\frac{dV(v)}{V_1}.$$

Aquí la integral es igual a

$$\int_{\Theta^\circ}\exp\{|x|ae_xv^T\}dV(v)/V,\quad V = \text{mes}\,\Theta^\circ,$$

donde Θ° es la superficie de una esfera unitaria, $e_x = x/|x|$. Por consiguiente, si designamos

$$\psi(t) = \int_{\Theta^\circ}\exp\{te_xv^T\}dV(v), \tag{12}$$

entonces, la región (11) de aceptación de H_2 tendrá la forma

$$\psi(|x|b) > c\psi(|x|a) \tag{13}$$

(aquí, por c designamos las constantes que no coinciden obligatoriamente con el valor en (11)). Pero, evidentemente, $\psi(t)$ no depende de x, puesto que el valor de la integral (12) no depende del sentido de dirección del vector unitario e_x. Por eso

$$\psi(t) = \int_{\Theta^\circ}\exp\{tv_1\}dV(v),$$

donde v_1 es la primera coordenada del vector v.

Como $\psi'(0) = 0$, $\psi''(t) > 0$ cuando $t > 0$, entonces $\psi(t)$ es una función convexa creciente en $[0, \infty)$. De aquí resulta que la desigualdad (13) u (11) equivale a

$$|x| > c. \tag{14}$$

Esto es, evidentemente, un criterio invariante. Comprobemos para él el cumplimiento de las condiciones 1—3 del teorema 1 y establezcamos asimismo que ello es el criterio minimax.

Tenemos

$$\mathbf{M}_\alpha \pi_{Q_1Q_2}(X) = \mathbf{P}_\alpha(|X| > c) = \Phi_{0,E}(\{x : |x - a| > c\}).$$

Está claro que el traslado del punto α en la esfera $|\alpha| = \text{const}$ no modifica dicha probabilidad. Por lo tanto, esta última sólo depende de $|\alpha|$ y, por consiguiente,

$$\mathbf{M}_\alpha \pi_{Q_1Q_2} = \mathbf{P}(|\xi - \alpha|^2 > c^2) =$$

$$= \mathbf{P}\left(\sum_{i=1}^{m} (\xi_i - \alpha_i)^2 > c^2 = \mathbf{P}((\xi_1 - |\alpha|)^2 + \xi_2^2 + \ldots + \xi_m^2 > c^2\right),$$

donde $\xi_i \in \Phi_{0,1}$ son las coordenadas independientes del vector ξ.

Lema 1. *La función* $p_c(t) = \mathbf{P}((\xi_1 - t)^2 + \xi_2^2 + \ldots + \xi_m^2 > c^2)$ *es para cada* c *la función creciente* $|t|$.

De este lema se desprende que

$$\mathbf{M}_\alpha \pi_{Q_1Q_2}(X) = p_c(|\alpha|) \leqslant p_c(a) \text{ cuando } |\alpha| \leqslant a,$$

$$\mathbf{M}_\alpha \pi_{Q_1Q_2}(X) = p_c(|\alpha|) \geqslant p_c(b) \text{ cuando } |\alpha| \geqslant b.$$

Estas relaciones equivalen a las condiciones 2) y 3) del teorema 1. Para que el criterio $\pi_{Q_1Q_2}$ sea el criterio de nivel $1 - \varepsilon$, debemos suponer que c es igual a la solución c_ε de la ecuación $p_c(a) = \varepsilon$. Ahora bien, $\pi_{Q_1Q_2}$ es el criterio minimax de nivel $1 - \varepsilon$ y su potencia garantizada es igual a $p_{c_\varepsilon}(b)$. ◁

Demostración del lema 1. Como $p_c(t) = p_c(-t)$, podemos limitarnos a examinar los valores de $t \geqslant 0$.

Examinemos primeramente el caso de $m = 1$. Designemos en este caso la función $p_c(t)$ por $p(t)$. Tenemos

$$p(t) = \mathbf{P}(|\xi_1 - t|^2 > c^2) = \Phi(t - c) + 1 - \Phi(t + c).$$

Por consiguiente, la derivada respecto a t es igual a

$$p'(t) = \frac{1}{\sqrt{2\pi}}[e^{-(t-c)^2/2} - e^{-(c+t)^2/2}] =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-(c^2+t^2)^2/2}[e^{ct} - e^{-ct}] \geqslant 0$$

y la función $p(t)$ crece cuando $t \geqslant 0$.

Cuando $m > 1$ la función $p_c(t)$ es la convolución de la función $p(t) = p(t, c^2)$ con la distribución χ^2 de $m - 1$ grados de libertad:

$$p_c(t) = \int_0^\infty p(t, c^2 - u)dH_{m-1}(u).$$

Evidentemente, ésta también es una función creciente de t para $t \geqslant 0$. ◁

En lo que se refiere al teorema 2 se puede señalar lo siguiente. Supongamos, para abreviar, que $a = 0$. Entonces, la hipótesis $H_1 = \{\alpha = 0\}$ será simple. Si construimos el c.m.p. para cada alternativa $\alpha \in \Theta_2$, obtendremos el criterio que tiene la forma

$$x_\alpha{}^T > c.$$

Esto significa que cada sentido de direccción de $\alpha = \alpha_0 t$, $\alpha_0 \in \Theta_2^\circ$, $t \geqslant 1$ tendrá su propio criterio más potente de nivel $1 - \varepsilon$

$$x\alpha_0^T > c_\xi, \tag{15}$$

donde c_ξ depende únicamente de ε, ya que $\mathbf{M}_0(X\alpha_0^T) = 0$. $\mathbf{D}_0(X\alpha_0^T) = |\alpha_0|^2 = b$. Pero la región crítica del criterio minimax (invariante) debe ser igualmente sensible respecto a todas las alternativas. En concordancia con esto, la misma tiene forma de unión de los semiespacios (15), que no es otra cosa sino el exterior de la esfera.

Ejemplo 2. Ahora supongamos que $X = x_1 \in \Phi_{\alpha,\sigma^2}$, donde $\sigma^2 = \|\sigma_{ij}\|$ es una matriz arbitraria de segundos momentos, definida positivamente. Examinemos el problema de verificación de la hipótesis $H_1 = \{\alpha\sigma^{-2}\alpha^T \leqslant a^2\} = \{|\alpha\sigma^{-1}| \leqslant a\}$ frente a $H_2 = \{\alpha\sigma^{-2}\alpha^T \geqslant b^2\} = \{|\alpha\sigma^{-1}| \geqslant b\}$, $a < b$. Del teorema 2 se deduce el

Teorema 2A. *El conjunto crítico del criterio minimax de nivel* $1 - \varepsilon$ *para verificar* H_1 *frente a* H_2 *tiene la forma*

$$x\sigma^{-2}x^T > c_\xi^2$$

y la potencia garantizada $p_{c_\xi}(b)$, donde c_ξ es, como antes, la solución de la ecuación $p_c(a) = \varepsilon$.

Demostración. Pongamos $gx = x\sigma$ y notemos que, en virtud de (7.3),

$$\Phi_{\alpha,E}(A) = \Phi_{\bar{g}(\alpha,E)}(gA),$$

donde $\bar{g}(\alpha, E) = (\alpha, \sigma^2)$. Para la esfera $A = \{x: |x| < c\}$ tendremos

$$\begin{gathered} gA = \{y = x\sigma: xx^T < c^2\} = \{y: y\sigma^{-2}y^T < c^2\}, \\ \Phi_{\alpha,E}(A) = \Phi_{\alpha,\sigma^2}(\{x: x\sigma^{-2}x^T < c^2\}). \end{gathered} \tag{16}$$

El conjunto $\{\alpha: |\alpha| \leqslant a\}$ pasa, después de la transformación $\bar{g}$, al conjunto $\{\beta = \alpha\sigma: \alpha\alpha^T \leqslant a^2\} = \{\beta: \beta\sigma^{-2}\beta^T \leqslant a^2\}$.

Ahora bien, todas las relaciones establecidas en el ejemplo 1 para $\Phi_{\alpha,E}(A)$ cuando $|\alpha| \leqslant a$ o cuando $|\alpha| \geqslant b$ serán válidas para $\Phi_{\beta,\sigma^2}(\{x: x\sigma^{-2}x^T < c^2\})$ cuando $|\beta_\sigma^{-1}| \leqslant a$ o bien $|\beta_\sigma^{-1}| \geqslant b$, respectivamente.

Esto demuestra el teorema 2A. ◁

Ejemplo 3. Volvamos a examinar la muestra de la distribución normal $\Phi_{\alpha,E}$ con una matriz unidad de segundos momentos. Sin embargo, a distinción del ejemplo 1, las hipótesis H_i sometidas a comprobación sólo tocarán una parte de las coordenadas del vector α. Representemos α en forma de un conjunto de dos vectores $\alpha = (\alpha', \alpha'')$, donde $\alpha' = \alpha_1, \ldots, \alpha_l)$, $\alpha'' = (\alpha_{l+1}, \ldots, \alpha_m)$, y examinemos el problema de verificación de la hipótesis $H_1 = \{|\alpha''| \leqslant a\}$ frente a $H_2 = \{|\alpha''| \geqslant b\}$, conforme a la muestra $X = \mathrm{x}_1 = (x_{1,1}, \ldots, x_{1,m})$ de volumen $n = 1$. Para cada una de las hipótesis, la magnitud α' puede adoptar un valor arbitrario. Procedamos del mismo modo que en el ejemplo 1, pero en calidad de $\mathbf{Q}_1$ y $\mathbf{Q}_2$ escojamos las distribuciones uniformes en las "esferas" $\Theta_1^\circ = \{\alpha: |\alpha''| = a, \alpha' = \alpha_0'\}$, $\Theta_2^\circ = \{\alpha: |\alpha''| = b, \alpha' = \alpha_0'\}$, donde α_0' es un punto registrado cualquiera. Si designamos $\mathrm{x}_1' = (x_{1,1}, \ldots, x_{1,l})$, $\mathrm{x}_1'' = (x_{1,l+1}, \ldots, x_{l,m})$, obtendremos como resultado el criterio minimax

$$|\mathrm{x}_1''| > c_\varepsilon,$$

donde c_ε es la solución de la ecuación

$$\mathbf{P}((\xi_1 - a)^2 + \xi_2^2 + \ldots + \xi_{m-l}^2 > c^2) = \varepsilon \tag{17}$$

(los factores $\exp\left\{-\frac{1}{2}(x' - \alpha_0')(x' - \alpha_0')^T\right\}$ en la desigualdad $f_{Q_2}(X) + f_{Q_1}(X) > c$ serán eliminados, y ésta se convertirá en una igualdad del tipo (11)). Este resultado es completamente natural, ya que en nuestro caso las coordenadas $x_{i,j}$ son independientes y, por lo tanto, el subvector x_1' no lleva en sí ninguna información respecto a α''. Por eso, de toda la muestra $X = \mathrm{x}_1$ sólo es suficiente examinar el subvector x_1'' y, en este caso, el problema se reduce al ejemplo 1.

La verificación de las hipótesis en el ejemplo 3 pertenece a la clase de problemas en que existe el llamado parámetro "obstaculizador". En nuestro caso, en calidad de tal parámetro servía el vector α'. En virtud de las causas mencionadas anteriormente, éste en realidad no obstaculizaba la construcción del criterio minimax, el cual automáticamente resultaba independiente de α'.

De manera algo diferente ocurre en el ejemplo siguiente, más general, cuando las coordenadas x_{ij} son dependientes.

Ejemplo 4. Supongamos que $X = x_1 \in \Phi_{\alpha,\sigma^2}$. Examinemos el problema de verificación de la hipótesis

$$H_1 = \{\alpha d^{-2}\alpha^T \leqslant a^2\} \text{ frente a } H_2 = \{\alpha d^{-2}\alpha^T \geqslant b^2\}, \tag{18}$$

donde d^{-2} es una matriz definida no negativamente de rango $m - l < m$, obtenida de σ^{-2} a base de sustituir por ceros los elementos de cualesquiera l renglones y l columnas (con los mismos números de orden). Para facilitar la exposición podemos considerar que, para la matriz definida positivamente σ_2^{-2} de orden $(m - l) \times (m - l)$, inversa a la matriz

$$\sigma_2^2 = \mathbf{M}_{\alpha,\sigma^2}(x_1'' - \alpha'')^T(x_1'' - \alpha''),$$

formada por las últimas $m - l$ columnas y renglones de la matriz $\sigma^2 = \|\sigma_{ij}\|$, se verifica la hipótesis $H_1 = \{\alpha''\sigma_2^{-2}\alpha''^T \leqslant a^2\}$ frente a $H_2 = \{\alpha''\sigma_2^{-2}\alpha''^T \geqslant b^2\}$, donde x_1'', α'' designan, al igual que en el ejemplo anterior, los mismos subvectores de los vectores x_1 y α. En cada una de las hipótesis H_i, el parámetro obstaculizador α' puede ser arbitrario.

Hablando en general, en este ejemplo, la distribución de x_1' depende de α''. Hagamos la siguiente transformación para convertir x_1 en vector con coordenadas "ortonormalizadas". Pongamos

$$Y = x_1\Lambda, \tag{19}$$

donde $\Lambda = \|a_{ij}\|$ es una matriz triangular con elementos $a_{ij} = 0$ $j > i$. Los restantes elementos se eligen de la condición $y \in \Phi_{\beta,E}$, donde $\beta = (\beta_1, \ldots$ $\ldots, \beta_m) = \alpha\Lambda$. Esto siempre se puede hacer, ya que de (19) obtenemos

$$y_m = x_{1,m}a_{m,m},$$
$$y_{m-1} = x_{1,m}a_{m,m-1} + x_{1,m-1}a_{m-1,m-1},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

De aquí y de las condiciones

$$\mathbf{M}_{\alpha,\sigma^2}(y_l - \beta_l)^2 = 1,$$
$$\mathbf{M}_{\alpha,\sigma^2}(y_l - \beta_i)(y_j - \beta_j) = 0, \ l \neq j,$$

se determinan uno tras otro los valores

$$a_{m,m}^2 = 1/\sigma_{m,m},$$
$$\sigma_{m,m}a_{m,m-1} + \sigma_{m-1,m}a_{m-1,m-1} = 0,$$
$$\sigma_{m,m}a_{m,m-1} + 2\sigma_{m,m-1}a_{m,m-1}a_{m-1,m-1} + \sigma_{m-1,m-1}a_{m-1,m-1}^2 = 1,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Ahora bien, la matriz triangular Λ es tal, que

$$\mathbf{M}_{\alpha,\sigma^2}(y-\beta)^T(y-\beta) = \mathbf{M}_{\alpha,\sigma^2}\Lambda^T(x_1-\alpha)^T(x_1-\alpha)\Lambda = \Lambda^T\sigma^2\Lambda = E.$$

Del carácter triangular de Λ se deduce que el vector $\beta'' = (\beta_{l+1}, \ldots, \beta_m)$

depende únicamente de α'', y al contrario. Si designamos por Λ_2 la matriz triangular de orden $(m - l) \times (m - l)$, obtenida de los últimos $m - l$ renglones y columnas de la matriz Λ entonces, obtenemos, evidentemente $\beta'' = \alpha''\Lambda_2$, $\Lambda_2^T\sigma_2^2\Lambda_2 = E$. El conjunto $\Theta_1 = \{\alpha: \alpha''\sigma_2^{-2}\alpha''^T \leqslant a^2\}$ se convertirá en el conjunto

$$\{\beta: \beta = \alpha\Lambda,\ \alpha''\sigma_2^{-2}\alpha''^T \leqslant a^2\} = \{\beta: \beta''\Lambda_2^{-1}\sigma_2^{-2}\Lambda_2^{-1T}\beta''^T \leqslant a^2\} =$$
$$= \{\beta: \beta''\beta''^T \leqslant a^2\} = \{\beta: |\beta| \leqslant a\}.$$

El "subparámetro" β' puede ser arbitrario si es arbitrario α'.

Hemos llegado al problema del ejemplo 3. El criterio minimax de nivel $1 - \varepsilon$ para verificar H_1 frente a H_2 tiene, por consiguiente, la forma $y''y''^T > c_\varepsilon$ o bien $(\Lambda_2\Lambda_2^T = \sigma_2^{-2})$

$$x_1''\sigma_2^{-2}x_1''^T > c_\varepsilon,$$

donde c_ε es la solución de la ecuación (17).

El último ejemplo es el más general entre los ejemplos 1—4. El mismo resume el contenido de estos ejemplos de la manera siguiente.

Teorema 2B. *Si a base de la muestra* $X = x_1 \in \Phi_{\alpha,\sigma^2}$ *se verifican las hipótesis* (18) *relacionadas con el valor* $\alpha, d^{-2}\alpha^T$, *entonces el criterio minimax de nivel* $1 - \varepsilon$ *tendrá la forma*

$$x_1 d^{-2} x_1^T > c_\varepsilon, \tag{20}$$

donde c_ε *se define en* (17), *y* $m - l$ *es el rango* d^{-2}.

La potencia garantizada del criterio (20) *es igual a*

$$\mathbf{P}((\xi_1 - b)^2 + \xi_2 + \ldots + y_{m-1} > c_\xi^2,\ \xi_i \in \Phi_{0,1}.$$

Si la muestra X tiene volumen n, entonces $\bar{x} \in \Phi_{\alpha,\sigma^2/n}$ tendrá la forma

$$\bar{x}d^{-2}\bar{x}^T > c_\varepsilon/n.$$

El siguiente ejemplo tiene, en cierta medida, otro carácter.

Ejemplo 5. Supongamos, al igual que en el ejemplo 1, que $X = x_1 \in \Phi_{\alpha,E}$ es una muestra de volumen $n = 1$ de una distribución normal m-dimensional de media $\alpha = (\alpha_1, \ldots, \alpha_m)$. Supongamos también, que $H_1 = \{\alpha = 0\}$ y que la hipótesis H_2 consiste en que α pertenece a cierto conjunto $\overline{\Theta}_2$ que no contiene los puntos $\alpha \in \Theta_2$. Designemos por $\overline{\Theta}_2$ la clausura convexa del conjunto Θ_2 (conjunto cerrado convexo mínimo que contiene Θ_2), y sea β el punto de $\overline{\Theta}_2$ más próximo al origen de coordenadas. Entonces, si $\beta \in \Theta_2$, la distribución $\mathbf{Q}_2$ concentrada en el punto β será la menos favorable, y el criterio minimax $\bar{\pi}$ tendrá la forma $\bar{\pi}(X) = 1$ si

$$(X - \beta)(X - \beta)^T < XX^T + c_1$$

o bien, que es lo mismo, si

$$X\beta^T/|\beta| > c_2,$$

donde c_2 se elige de la condición $\bar{\pi} \in K_\varepsilon$.

En efecto, es suficiente comprobar la condición (7). Tenemos

$$\mathbf{M}_\alpha \bar{\pi}(X) = \mathbf{P}_\alpha(X\beta^T/|\beta| > c_2),$$

donde $X\beta^T/|\beta| \in \Phi_{\alpha\beta^T/|\beta|,1}$, así que

$$\mathbf{M}_\alpha \bar{\pi}(X) = 1 - \Phi(c_2 - \alpha\beta^T/|\beta|).$$

Esto significa que el mínimo $\mathbf{M}_\alpha \bar{\pi}(X)$, $\alpha \in \Theta_2$ se alcanza para α que minimiza la función $\alpha\beta^T/|\beta|$. Pero es evidente que $\alpha\beta^T \geqslant \beta\beta^T = |\beta|^2$ para todos $\alpha \in \Theta_2$, así que

$$\mathbf{M}_\beta \bar{\pi}(X) = \inf_{\alpha \in \Theta_2} \mathbf{M}_\alpha \bar{\pi}(X). \quad \triangleleft$$

Le proponemos al lector que construya el criterio minimax conforme a ese mismo problema, es decir, cuando $X \in \Phi_{\alpha,\sigma^2}$, σ^2 es una matriz arbitraria de segundos momentos.

3. Distribuciones degeneradas menos favorables para las hipótesis unilaterales. Supongamos que $X \in \mathbf{P}_\theta$, donde θ y los elementos x_i de la muestra X son reales.

Supongamos además, que verificamos la hipótesis unilateral $H_1 = \{\theta \leqslant \theta_1\}$ frente a $H_2 = \{\theta \geqslant \theta_2\}$ siempre que haya una "zona de indiferencia" no vacía $\theta_1 < \theta < \theta_2$. ¿A qué condiciones las distribuciones menos favorables quedarán concentradas en los puntos θ_1 y θ_2? Pues en este caso el criterio minimax $\bar{\pi}$ de nivel $1 - \varepsilon$ tendría una forma muy simple:

$$\bar{\pi}(X) = \begin{cases} 1, & \text{si } f_{\theta_2}(X) > cf_{\theta_1}(X) \\ p, & \text{si } f_{\theta_2}(x) = cf_{\theta_1}(X), \\ 0, & \text{si } f_{\theta_2}(X) < cf_{\theta_1}(X), \end{cases} \tag{21}$$

donde p y c se definen por la igualdad $\mathbf{M}_{\theta_1} \bar{\pi}(X) = \varepsilon$.

Ya sabemos que si la relación de verosimilitud es monótona, tal criterio será el c.u.m.p. y, por consiguiente, también será minimax. La siguiente afirmación ofrece otra condición suficiente para que el criterio sea minimax.

Teorema 3. *Supongamos que la densidad $f_\theta(x)$ posee la propiedad de que la relación $f_{\theta'}(x)/f_\theta(x)$ no decrece respecto a x para cualesquiera $\theta' > \theta$. Entonces las distribución* $\mathbf{Q}_1$ y $\mathbf{Q}_2$ *menos favorables estarán concentradas en los puntos θ_1 y θ_2, respectivamente, y, por lo tanto, el criterio* (21) *será minimax.*

Demostración. Supongamos primeramente que $n = 1$. Según las condiciones del teorema, habrá $a \leqslant b$ tales, que $f_{\theta'}(x)/f_\theta(x) \leqslant 1$ cuando $x \in (-\infty, a]$, $f_{\theta'}(x)/f_\theta(x) = 1$ cuando $x \in (a, b)$ y $f_{\theta'}(x)/f_\theta(x) \geqslant 1$ cuando $x \in [b, \infty)$. Como $\bar{\pi}(x)$ no decrece, entonces $\bar{\pi}(b) \geqslant \bar{\pi}(a)$ y

$$\mathbf{M}_{\theta'}\bar{\pi}(X) - \mathbf{M}_\theta\bar{\pi}(X) \geqslant$$

$$\geqslant \bar{\pi}(a) \int_{-\infty}^{a} (f_{\theta'}(x) - f_\theta(x))\mu(dx) + \bar{\pi}(b) \int_{b}^{\infty} (f_{\theta'}(x) - f_\theta(x))\mu(dx) =$$

$$= (\bar{\pi}(b) - \bar{\pi}(a)) \int_{b}^{\infty} (f_{\theta'}(x) - f_\theta(x))\mu(dx) \geqslant 0.$$

Si $n > 1$, para obtener esta misma desigualdad es necesario valerse de la integración sucesiva (primero respecto a x_1, luego respecto a x_2, etc.) y del hecho de que $\bar{\pi}(X)$ no decrece con arreglo a cada uno de sus argumentos.

Ahora bien, hemos establecido que la potencia $\beta(\theta) = \mathbf{M}_\theta\pi(X)$ es una función no decreciente.

De aquí se deduce que el nivel de $\bar{\pi}$ es igual a $1 - \varepsilon$ y que $\beta(\theta_1) = \sup_{\theta < \theta_1} \beta(\theta)$ y $\beta(\theta_2) = \inf_{\theta > \theta_2} \beta(\theta)$. Esto significa que se cumplen todas las condiciones del teorema 1. El teorema 3 queda demostado. $\triangleleft$

Si θ es el parámetro de desplazamiento: $f_\theta x) = f(x - \theta)$, se puede mostrar que $f_{\theta'}(x)/f_\theta(x)$ será monótona respecto a x si y sólo si la función $-\ln f(x)$ es convexa (véase [57]).

§ 10. Criterio de la relación de verosimilitud

En los párrafos anteriores hemos obtenido varios resultados concernientes a la construcción de todo género de criterios óptimos. Una deducción importante que se puede sacar de las consideraciones citadas consiste en que estos criterios óptimos sólo existen en condiciones bastante limitadas. En la teoría de la estimación hemos tenido, aproximadamente, la misma situación: las estimaciones eficientes también existen únicamente en condiciones limitadas. No obstante, en el capítulo 2 hemos visto que si se examina no la propiedad exacta de eficacia, sino la propiedad asintótica, entonces las estimaciones que poseen esta propiedad ya existen muy a menudo en condiciones relativamente amplias, relacionadas casi siempre con la regularidad de la familia $\{\mathbf{P}_\theta\}$. Tales condiciones son las e.v.m.

Otra expresión de la optimización asintótica de la e.v.m. consiste como hemos visto, en que las e.v.m. son asintóticamente equivalentes a las estimaciones bayesianas para cualquier distribución a priori suave registrada.

En la teoría de verificación de las hipótesis, cierto análogo de la e.v.m. es el llamado *criterio de la relación de verosimilitud* (c.r.v.). En caso de

amplias suposiciones, el referido criterio coincide con los criterios óptimos, si tales existen, y resulta asintóticamente equivalente al criterio bayesiano cuando $\Theta_1 = \{\theta_1\}$ para cualquier distribución a priori suave registrada Q_2 en Θ_2. Esta propiedad y una serie de otras propiedades asintóticas del c.r.v. serán establecidas en los párrafos inmediatos.

Demos la definición del c.r.v. Supongamos que en el caso paramétrico, cuando $X \in \mathbf{P}_\theta$, se verifica la hipótesis $H_1 = \{\theta \in \Theta_1\}$ frente a la hipótesis $H_2 = \{\theta \in \Theta_2\}$.

Definición 1. El criterio $\hat{\pi}(X)$ con la región crítica

$$R(X) = \frac{\sup\limits_{\theta\in\Theta_2} f_\theta(X)}{\sup\limits_{\theta\in\Theta_1} f_\theta(X)} > c \tag{1}$$

se llama *criterio de la relación de verosimilitud* (c.r.v.) para verificar la hipótesis H_1 frente a H_2.

La constante c suele elegirse de la condición

$$\sup_{\theta\in\Theta_1} \mathbf{P}_\theta(R(X) > c) = \varepsilon, \tag{2}$$

para la cual el c.r.v. tendrá un nivel de $1 - \varepsilon$.

A la par con el criterio (1) a menudo se examina un criterio que, de hecho, equivale al primero (también llamado c.r.v) y que tiene la forma siguiente:

$$R_1(X) = \frac{\sup\limits_{\theta\in\Theta} f_\theta(X)}{\sup\limits_{\theta\in\Theta_1} f_\theta(X)} = \frac{f_{\theta^*}(X)}{\sup\limits_{\theta\in\Theta_1} f_\theta(X)} > c. \tag{3}$$

La semejanza de estos criterios se desprende del hecho de que cuando $\Theta = \Theta_1 \cup \Theta_2$,

$$f_{\theta^*}(X) = \max\{\sup_{\theta\subset\Theta_1} f_\theta(X),\ \sup_{\theta\in\Theta_2} f_\theta(X\}$$

y, por lo tanto, $R_1(X) = \max\{1,\ R(X)\}$.

Si la hipótesis H_1 es simple: $\Theta_1 = \{\theta_1\}$, $H_2 = \{\theta \neq \theta_1\}$, así que $\Theta_2 = \Theta \setminus \{\theta\}$, entonces para $f_\theta(x)$, continuas respecto a θ, tendremos

$$R(X) = R_1(X) = f_{\theta^*}(x)/f_{\theta_1}(X).$$

Según su forma, el criterio (1) generaliza de un modo natural el c.m.p. para verificar las hipótesis simples en el lema de Neumann—Pearson. Y aunque en el caso general este criterio no tiene, por lo visto, *exactas* propiedades de optimización, a menudo resulta ser el mejor asintóticamente (véanse los §§ 13—16).

Muchos criterios invariantes y minimax no desplazados, examinados más arriba, son los c.r.v. En calidad de ilustración examinemos los ejemplos

9.1—9.4 donde se construyeron los criterios minimax para el parámetro α de poblaciones normales. *En todos estos ejemplos, los criterios minimax son los c.r.v.* Demostrémoslo. Los problemas de los ejemplos 9.2 y 9.4 se han reducido, con una exactitud de hasta las transformaciones lineales del parámetro, a los problemas de los ejemplos 9.1 y 9.3. En vista de que la relación de verosimilitud (1) no depende de tales sustituciones (al variar respectivamente las regiones Θ_i), es suficiente examinar tan sólo los ejemplos 9.1 y 9.3.

En el ejemplo 9.1, a base de una muestra $X \in \Phi_{\alpha,E}$ de volumen unitario y procedente de una población normal multidimensional con una matriz unidad E de segundos momentos, hemos verificado la hipótesis $H_1 = \{|\alpha| \leqslant a\}$ frente a $H_2 = \{|\alpha| \geqslant b\}$, $a < b$. Resultó que el criterio minimax tiene la forma

$$|X| > c. \tag{4}$$

En nuestro caso, $\sup\limits_{\theta\in\Theta_i} f_\theta(X)$ se define por el valor

$$\inf_{\alpha\in\Theta_i} (X - \alpha)(X - \alpha)^T = \inf_{\theta\in\Theta_2} |X - \alpha|^2,$$

así que para la estadística $R(X)$ en (1) tendremos

$$\ln R(X) = \begin{cases} -\frac{1}{2}(|X| - b)^2, & \text{si} |X| \leqslant a, \\ -\frac{1}{2}(|X| - b)^2 + \frac{1}{2}(|X| - a)^2, & \text{si } a < |X| < b, \\ -\frac{1}{2}(|X| - a)^2, & \text{si } |X| \geqslant b. \end{cases}$$

Esta es una función creciente continua de $|X|$. Por eso las regiones (1) y (4) coinciden para valores convenientes de c.

Le proponemos al lector que él mismo se cerciore de que en este ejemplo el criterio (3) también tiene la forma (4).

En el ejemplo 9.3, a base de la muestra $X \in \Phi_{\alpha,E}$ de volumen unitario, hemos verificado la hipótesis $H_1 = \{|\alpha''| \leqslant a\}$ frente a $H_2 = \{|\alpha''| \geqslant b\}$, donde $\alpha'' = (\alpha_{l+1}, \ldots, \alpha_m)$ es un subvector del vector α constituido por sus últimas $m - l$ coordenadas. El criterio minimax tiene la forma

$$|X''| > c, \tag{5}$$

donde X'' está constituido por las últimas $m - l$ coordenadas del vector X. Pero en este caso

$$\inf_{\alpha\in\Theta_1} (X - \alpha)(X - \alpha)^T = \inf_{\alpha'':|\alpha''|\leqslant a} (X'' - \alpha'')(X'' - \alpha'')^T.$$

La desigualdad análoga es válida para Θ_2. Por eso todo se reduce a las consideraciones del ejemplo 9.1, y los c.r.v. (1) y (3) coincidirán con (5).

En condiciones del § 5, los c.u.m.p. allí construidos para las familias exponenciales

$$f_\theta(x) = c(\theta)e^{\theta T(x)}h(x) \tag{6}$$

también coincidirán con los c.r.v. El lector puede comprobar esto personalmente, notando que la función

$$\varphi(\theta) = \ln c(\theta) = -\ln\left(\int e^{\theta T(x)}h(x)\mu^n(dx)\right)$$

es convexa, puesto que $\varphi'(\theta) = -\mathbf{M}_\theta T$, $\varphi''(\theta) = -\mathbf{D}_\theta T < 0$. De la convexidad de φ se deduce la solubilidad unívoca de la ecuación

$$\varphi'(\theta) + T(X) = 0$$

para la e.v.m. $\hat{\theta}^* = \psi(T)$ y la monotonía de la función φ. En este caso, uno de los $\sup_{\theta\in\Theta_i} f_\theta(X)$ se alcanzará en el punto $\hat{\theta}^*$, y el otro, en los puntos θ_1 o θ_2.

La verificación de la referida afirmación para las familias normales $\Phi_{\alpha,E}$, que son un caso particular de (6), se expone en el § 15.

Es algo diferente el asunto examinado en el ejemplo 9.5, donde, de acuerdo con la muestra $X \in \Phi_{\alpha,E}$ hemos verificado la hipótesis $H_1 = \{\alpha = 0\}$ frente a $H_2 = \{\alpha \in \Theta_2\}$. Se supone que el conjunto Θ_2 y su clausura convexa $\overline{\Theta}_2$ no contienen puntos $\alpha = 0$. Si el punto β más próximo al origen de coordenadas del conjunto $\overline{\Theta}_2$ pertenece a Θ_2, entonces el criterio minimax existe y tiene la forma siguiente:

$$X\beta^T > c. \tag{7}$$

Este criterio no es invariante respecto a cualquier grupo de transformaciones. Le proponemos al lector que él mismo se cerciore de que en este caso el c.r.v. es distinto de (7) y tiene la forma

$$\varrho^2(X, \Theta_2) - \varrho^2(X, 0) < c,$$

donde $\varrho(X, \Theta_2) = \inf_{\alpha\in\Theta_2} |X - \alpha|$, $\varrho(X, 0) = |X|$.

Ahora demostraremos que cuando se cumplen ciertas suposiciones, el criterio de la relación de verosimilitud posee propiedades de invariación. Sea G cualquier grupo de transformaciones en $\mathscr{X}^n$, respecto al cual el problema de verificación de las hipótesis H_1 y H_2 es invariante, y sea $\overline{G}$ el grupo respectivo de transformaciones $\bar{g}$ en Θ.

Teorema 1. *Si $f_\theta(x)$ posee la propiedad*

$$f_\theta(gx) = c(g, x)f_{\bar{g}\theta}(x), \tag{8}$$

entonces el criterio de la relación de verosimilitud es invariante respecto a G.

En cuanto a la condición (8) diremos que la misma siempre se cumple cuando μ es la medida de Lebesgue, y g, la transformación que conserva esa medida (desplazamiento y giro). En este caso $c(g, x) \equiv 1$. Para las transformaciones de contracción, $c(g, x) \equiv$ const.

Demostración del teorema 1. En virtud de que $\bar{g}\Theta_i = \Theta_i$, $i = 1, 2$, tendremos

$$R(gx) = \frac{\sup\limits_{\theta\in\Theta_2} f_\theta(gx)}{\sup\limits_{\theta\in\Theta_1} f_\theta(gx)} = \frac{\sup\limits_{\theta\in\Theta_2} c(g, x) f_{\bar{g}\theta}(x)}{\sup\limits_{\theta\in\Theta_1} c(g, x) f_{\bar{g}\theta}(x)} = \frac{\sup\limits_{\theta\in\bar{g}\Theta_2} f_\theta(x)}{\sup\limits_{\theta\in\bar{g}\Theta_1} f_\theta(x)} = R(x). \quad \triangleleft$$

Otras propiedades del c.r.v. véanse en los §§ 11, 13—16.

§ 11*. Análisis sucesivo

1. Observaciones preliminares. En todos los planteamientos anteriores, el volumen n de la muestra $X = X_n$, de la cual disponemos, estaba registrado. En tales condiciones hemos hallado criterios que poseían unas u otras propiedades de optimización. Por ejemplo, en el caso más elemental, cuando se verificaban dos hipótesis simples $H_i = \{X \in \mathbf{P}_i\}$, $i = 1, 2$, resultó que existe un c.m.p. π de nivel $1 - \varepsilon$, el cual tiene la forma (véase el teorema 2.1)

$$\pi(X) = \begin{cases} 1, & \text{si } f_2(X) > cf_1(X), \\ p, & \text{si } f_2(X) = cf_1(X), \\ 0, & \text{si } f_2(X) < cf_1(X). \end{cases}$$

Aquí c y p se deducen de la condición $\mathbf{M}_1\pi(X) = \varepsilon$, y $f_i(x)$ son las densidades de las distribuciones $\mathbf{P}_i$, $i = 1, 2$, respecto a cierta medida μ.

¿Será posible mejorar ulteriormente este procedimiento estadístico? En las condiciones enunciadas claro está que no es posible. Pero si desistimos en registrar el volumen de la muestra, o sea, si procedemos a que el número de observaciones n sea una variable aleatoria dependiente de las observaciones ya realizadas, entonces los mejoramientos son posibles. Se tiene en cuenta la reducción de la cantidad de observaciones indispensables para construir los criterios a base de ciertos parámetros dados. Esta circunstancia es importante en los experimentos donde la ejecución de ensayos ofrece gastos considerables.

La posibilidad de tal mejoramiento de los criterios puede ser aclarada citando el ejemplo siguiente. Supongamos que las distribuciones $\mathbf{P}_1$ y $\mathbf{P}_2$ no son del todo recíprocamente continuas, y supongamos también, que existen conjuntos B_1 y B_2 de $\mathfrak{B}_{\mathscr{X}}$ tales, que $f_1(x) > 0$, $f_2(x) = 0$ cuando $x \in B_1$, y $f_1(x) = 0$, $f_2(x) > 0$ cuando $x \in B_2$. Entonces está claro que si

$x_1 \in B_1$ ($x_1 \in B_2$), podemos afirmar infaliblemente que tiene lugar la hipótesis H_1 (H_2). En este caso no hay ninguna necesidad de llevar a efecto las observaciones posteriores.

Ahora bien, si los experimentos se realizan no de una vez (en cantidad de n), sino sucesivamente, examinando el resultado de cada nueva serie de observaciones, entonces es posible reducir el volumen general de observaciones.

La introducción del procedimiento sucesivo también es muy natural desde el punto de vista del enfoque bayesiano. En efecto, el referido enfoque, examinado en el § 2, prescribe aceptar la hipótesis H_2 si la probabilidad a posteriori $q(2/X)$ de esta hipótesis $\geqslant 1/2$. En este caso, en el conjunto crítico se encontrarán, entre otras, tanto muestras X para las cuales $q(2/x)$ es próxima a 1 (para tales X, la aceptación de H_2 es oportuna), como muestras X para las cuales $q(2/X)$ es próxima a 1/2. Estas últimas podrían considerarse como muestras "insuficientes" para tomar decisiones y las cuales requieren experimentos adicionales. Además, al igual que en el ejemplo expuesto más arriba, la probabilidad a posteriori $q(2/X)$ puede resultar grande ya después de las primeras pruebas, y entonces se podría tomar decisiones sin efectuar pruebas posteriores (en el ejemplo mencionado, $q(2/X) = 1$ cuando $X = x_1 \in B_2$ para cualquier distribución a priori $(q(1), q(2))$, $q(2) > 0$).

Más abajo examinaremos el procedimiento sucesivo para verificar dos hipótesis simples, en el cual se alcanzará la reducción máxima posible de la cantidad de observaciones.

2. Criterio sucesivo bayesiano. Examinemos primeramente el planteamiento bayesiano del problema y designemos por $q(1) = q$ y $q(2) = 1 - q$ las probabilidades a priori de las hipótesis H_1 y H_2. Entonces, la probabilidad a posteriori de la hipótesis H_i después de las observaciones $X = X_n$ será igual a

$$q(i/X_n) = \frac{q(i)f_i(X_n)}{q(1)f_1(X_n) + q(2)f_2(X_n)}. \quad (1)$$

Realizaremos sucesivamente las observaciones y para cada n calcularemos los valores de $q(2/X_n)$, $n = 1, 2, \ldots$ (o de $q(1/X_n)$). En el plano de las variables (n, y) examinaremos la trayectoria aleatoria de las probabilidades a posteriori (quebrada aleatoria), que parte del punto $q = q(2)$ cuando $n = 0$ y que toma, en los puntos $n = 1, 2, \ldots$, los valores de $y = q(2/X_n)$. Con ayuda de esta trayectoria se puede construir el siguiente criterio para verificar la hipótesis H_1 frente a H_2: examinemos en el plano (n, y) dos fronteras rectilíneas $y = \gamma_i$, $i = 1, 2$; $0 < \gamma_1 < \gamma_2 < 1$ para la variable $q(2/X_n)$. Se acepta la hipótesis H_2 si la trayectoria $q(2/X_n)$, $n = 0, 1, \ldots$, sale por primera vez de la franja (γ_1, γ_2) a través de la frontera superior γ_2. Si la trayectoria $q(2/X_n)$, $n = 0, 1, \ldots$, sale de esta franja a través de

la frontera inferior γ_1, entonces se acepta H_1. Más adelante veremos que la $\mathbf{P}_i$-probabilidad ($i = 1, 2$) de que $q(2/X_n)$ nunca saldrá de la franja (γ_1, γ_2), o sea, la probabilidad del suceso

$$\{\gamma_1 < q(2/X_n) < \gamma_2,\ n = 0, 1, \ldots\}$$

es igual a cero.

El número de pruebas ν que se necesita para aceptar una de las hipótesis (o sea, para alterar las desigualdades (2)) es, evidentemente, *variable aleatoria markoviana* (momento de parada) respecto a la sucesión $x_1, x_2, \ldots$ para cada una de las distribuciones $\mathbf{P}_1$ y $\mathbf{P}_2$. Desde este punto de vista, dicha regla de aceptación de las hipótesis es *sucesiva* y concuerda bastante bien con las reglas conforme a las cuales actúa el hombre en su actividad práctica: tomar una u otra decisión después que las observaciones permitan reducir en sumo grado la incertidumbre que tiene lugar con respecto al objeto sometido a examen.

El criterio construido depende de $q = q(1)$ y del vector $\gamma = (\gamma_1, \gamma_2)$. Por eso, designémoslo por $\delta_{q,\gamma}$. Ahora establezcamos que el criterio $\delta_{q,\gamma}$ es óptimo. Con este fin introduzcamos primeramente el concepto general de criterio sucesivo, cuyas características esenciales, a la par con las probabilidades de los errores de primero y segundo genero, se convierten en los valores medios $\mathbf{M}_1\nu$ y $\mathbf{M}_2\nu$ para el número de observaciones ν necesarias para tomar decisiones.

Supongamos que en $(\mathscr{X}^\infty, \mathfrak{B}_{\mathscr{X}}^\infty)$ se da una variable aleatoria entera arbitraria $\nu \geqslant 0$ que es markoviana respecto a la sucesión $x_1, x_2, \ldots$ ($\{\nu \geqslant n\} \in \sigma(x_1, \ldots, x_n) = \mathfrak{B}_{\mathscr{X}}^n$). Designemos por $\mathscr{X}^\nu$ el espacio de los vectores (n, X_n) tales, que $\nu(X\infty) = n$, $X_n = [X_\infty]_n$. Introduzcamos en $\mathscr{X}^\nu$ la σ-álgebra de $\mathfrak{B}^\nu$ engendrada por los sucesos $\{\nu = n, X_n \in B^n\}$, $B^n \in \mathfrak{B}_{\mathscr{X}}^n$, $n = 0, 1, \ldots$. Está claro que cualquier distribución en $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}})$ (o en $(\mathscr{X}^\infty, \mathfrak{B}_{\mathscr{X}}^\infty)$) induce la distribución respectiva en $(\mathscr{X}^\nu, \mathfrak{B}^\nu)$.

Definición 1. Llámase *criterio sucesivo* δ para verificar H_1 frente a H_2, el par (ν, Ω), donde $\Omega \in \mathfrak{B}^\nu$ es la región de aceptación de H_2 (región crítica), y la variable aleatoria ν se supone que es propia respecto a ambas distribuciones $\mathbf{P}_1$, $\mathbf{P}_2$ ($\mathbf{P}_i(\nu < \infty) = 1$, $i = 1, 2$).

En los casos cuando sea necesario señalar que ν y Ω pertenecen al criterio δ, escribiremos $\nu(\delta)$ y $\Omega(\delta)$.

Es natural que, de un modo equivalente, el criterio sucesivo puede ser designado con ayuda de una función biforme medible en $\mathscr{X}^\nu$. También está claro que el criterio sucesivo δ puede ser designado mediante la construcción de la región crítica (volvamos a designarla por Ω) en todo el espacio $\mathscr{X}^\infty$. Sin embargo, con tal aplicación (en $\mathscr{X}^\infty$) de las regiones Ω y $\mathscr{X}^\nu \setminus \Omega$ de aceptación de las hipótesis H_2 y H_1, no obtendremos obligatoriamente todos los elementos de $\mathscr{X}^\infty$: en aquellos de ellos para los cuales $\nu(X_\infty) = \infty$,

no se acepta ninguna hipótesis. Pero según la definición de la $\mathbf{P}_i$-probabilidad, los conjuntos de tales X_∞ equivalen a cero.

El criterio no randomizado ordinario δ es un caso particular del criterio sucesivo, cuando $\nu(\delta) \equiv n$ es constante (si $\nu(\delta) \equiv 0$, entonces la decisión se toma sin realizar ensayos).

El criterio sucesivo δ, al igual que cualquier criterio ordinario para verificar dos hipótesis simples, se caracteriza por las probabilidades $\alpha_i(\delta)$ de errores de i-ésimo género ($i = 1, 2$):

$$\alpha_i(\delta) = \mathbf{P}_i((\nu, X_\nu) \notin \Omega_i),$$

donde $\Omega_2 = \Omega$, $\Omega_1 = \mathscr{X}^\nu \setminus \Omega_2$. Además, como ya hemos señalado, caracterizaremos el criterio sucesivo por los valores medios $\mathbf{M}_i\nu$, $i = 1, 2$. Es evidente que para el criterio ordinario δ, construido según la muestra X_n, se cumple $\mathbf{M}_i\nu(\delta) \equiv n$.

Para tomar en consideración la aparición de estos nuevos factores en el planteamiento del problema (o sea, de las características relacionadas con la magnitud ν), supondremos que la realización de cada observación necesita gastos de valor a. También será cómodo caracterizar las pérdidas que surgen al tomar decisiones incorrectas, por medio de distintos valores de w_1 y w_2. Es decir, consideraremos que las pérdidas de i-ésimo género que surgen al tomar decisiones erróneas, cuando es cierta H_i, equivalen a w_i, $i = 1, 2$.

Con estos acuerdos, la esperanza matemática $R(q, \delta)$ de las pérdidas que surgen al utilizar el criterio δ, es igual a

$$R(q, \delta) = q[\alpha_1(\delta)w_1 + a\mathbf{M}_1\nu(\delta)] + (1 - q)[\alpha_2(\delta)w_2 + a\mathbf{M}_2\nu(\delta)]. \quad (3)$$

Esta expresión se denomina *riesgo bayesiano* en el problema sujeto a examen. Si aquí suponemos que $a = 0$, $w_1 = w_2 = 1$, obtendremos la expresión para la probabilidad de una decisión errónea del criterio δ, la cual ya hemos utilizado repetidas veces en los §§ 1, 2.

Definición 2. El criterio sucesivo δ que minimiza el riesgo bayesiano (3) se denomina criterio sucesivo *bayesiano*.

La siguiente afirmación establece la optimización (carácter bayesiano) del criterio $\delta_{a,\gamma}$ construido al principio de este párrafo.

Teorema 1. *Para a, w_1, w_2 dados existen γ_1, γ_2 tales, que el criterio $\delta_{a,\gamma}$ es bayesiano.*

Demostración. Designemos por δ_i el criterio que acepta la hipótesis H_i sin realizar pruebas, así que $\nu(\delta_i) = 0$, $\alpha_i(\delta_i) = 0$. Aclaremos primeramente en qué casos el criterio δ, que minimiza $R(q, \delta)$, coincide con δ_1 o con δ_2. Es evidente que

$$R(q, \delta_1) = (1 - q)w_2, \quad R(q, \delta_2) = qw_1.$$

Sea K la clase de criterios $\{\delta = \delta(X)\}$ que dependen al menos de una observación, o sea, la clase de criterios δ para los cuales $\nu(\delta) \geqslant 1$. Es evidente que $R(q, \delta) \geqslant a$ para $\delta \in K$. Designemos

$$R(q) = \inf_{\delta \in K} R(q, \delta).$$

Como el criterio δ, basado en una sola prueba ($\nu(\delta) = 1$), pertenece a K, entonces $R(q) < \infty$.

Para cualquier $p \in (0, 1)$ tenemos, en virtud de la linealidad de $R(q, \delta)$ como función de q:

$$R(pq_1 + (1 - p)q_2) = \inf_{\delta \in K} [pR(q_1, \delta) + (1 - p)R(q_2, \delta)] \geqslant$$

$$\geqslant pR(q_1) + (1 - p)R(q_2).$$

Esto quiere decir que $R(q)$ es una función cóncava. En vista de que $a < R(q) < \infty$, de aquí se deduce que $R(q)$ también es una función continua en $[0, 1]$. Comparemos ahora los riesgos de los criterios δ_i y $\delta \in K$ en función de q (véase la fig. 5).

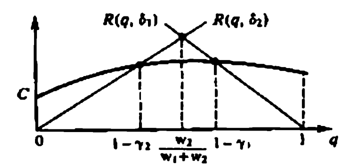


Fig. 5.

Una de dos: o bien $R(q) \geqslant \min R(q, \delta_i)$ para todos q (esto corresponde al hecho de que $R\left(\frac{w_2}{w_1 + w_2}\right) \geqslant \frac{w_1 w_2}{w_1 + w_2}$, o bien existen soluciones de las ecuaciones $R(q, \delta_1) = R(q)$, $R(q, \delta_2) = R(q)$, que designaremos $1 - \gamma_1$, $1 - \gamma_2$, $1 - \gamma_1 > 1 - \gamma_2$, respectivamente. Es evidente que $R(q) < \min R(q, \delta_i)$ dentro del intervalo $(1 - \gamma_2, 1 - \gamma_1)$. Para la primera de las posibilidades mencionadas supongamos

$$1 - \gamma_1 = 1 - \gamma_2 = \frac{w_1}{w_1 + w_2},$$

así que

$$R(1 - \gamma_1, \delta_1) = R(1 - \gamma_1, \delta_2).$$

De los referidos razonamientos y de la fig. 5 se deduce la siguiente regla óptima de acciones. A base de los datos a, w_1, w_2 calculamos $1 - \gamma_1$, $1 - \gamma_2$. Si $q \leqslant 1 - \gamma_2$ o bien, que es lo mismo, $1 - q \geqslant \gamma_2$, el menor riesgo

entre todos los criterios lo proporciona δ_2 (o sea, es necesario aceptar inmediatamente H_2). Si $q \geqslant 1 - \gamma_1$ $(1 - q \leqslant \gamma_1)$, entonces δ_1 ofrece el menor riesgo (es preciso aceptar H_1). Y sólo en el caso de $1 - \gamma_2 < 1 - \gamma_1$, $q \in (1 - \gamma_2, 1 - \gamma_1)$ (o bien $1 - q \in (\gamma_1, \gamma_2)$) es necesario utilizar el criterio de K, o sea, hay que realizar el experimento.

Ahora aprovechemos la inducción. Supongamos que se han efectuado n observaciones y que disponemos de la muestra X_n. Antes de la observación $n + 1$ tenemos la misma alternativa: no realizar más observaciones y aceptar una de las hipótesis H_i, o bien continuar las observaciones. El hecho de que ya hemos sufrido las pérdidas an no desempeña ningún papel, ya que éstas no pueden ser eliminadas de ningún modo. Los cambios esenciales sólo están relacionados con la distribución a priori. Ahora el papel de probabilidades $q(1) = q$ y $q(2) = 1 - q$ deben desempeñarlo las probabilidades a posteriori $q(1/X_n)$, $q(2/X_n)$. Con arreglo a esta nueva situación, la regla óptima ya elaborada por nosotros, dice que es necesario aceptar H_2 si $q(2/X_n) \geqslant \gamma_2$, y H_1 si $q(2/X_n) \leqslant \gamma_1$. Si $q(2/X_n) \in (\gamma_1, \gamma_2)$, entonces conviene continuar las observaciones. Pero la regla obtenida no es otra cosa sino el criterio $\delta_{q,\gamma}$. Ahora bien, hemos hallado $\gamma_i = \gamma_i(a, w_1, w_2)$ que poseen la propiedad de que el criterio $\delta_{q,\gamma}$ minimiza el riesgo $R(q, \delta)$. ◁

Nótese que los números $\gamma_i(a, w_1, w_2)$ permanecen invariables al multiplicar a, w_1, w_2 por un mismo número: esto es evidente de su definición, ya que tal operación sólo conduce a que todos los riesgos $R(q, \delta)$ sean multiplicados por ese mismo número. Así pues, en realidad γ_i es una función de dos variables, por ejemplo, de a y w_1 si consideramos que $w_2 = 1 - w_1$.

¿Qué representa en sí el criterio bayesiano $\delta_{q,\gamma}$? El mismo prescribe no realizar observaciones en dos casos: cuando $\gamma_1 = \gamma_2$ (lo cual sucede en caso de que a es grande en comparación con w_1, w_2), o bien cuando $q(2) \leqslant \gamma_1$ o cuando $q(2) \geqslant \gamma_2$. En los demás casos es preciso realizar experimentos hasta la primera alteración de las desigualdades

$$\gamma_1 < q(2/X_n) < \gamma_2$$

o bien, que es lo mismo, (véase (1)), hasta la primera alteración de las desigualdades

$$\frac{\gamma_1 q(1)}{(1 - \gamma_1)q(2)} < \frac{f_2(X_n)}{f_1(X_n)} < \frac{\gamma_2 q(1)}{(1 - \gamma_2)q(2)}. \tag{4}$$

En este caso se acepta la hipótesis H_2 si por primera vez se altera la desigualdad derecha, y la hipótesis H_1 si se altera la desigualdad izquierda. En tal forma, la parte "variable" del criterio $\delta_{q,\gamma}$ ya no está relacionada con el planteamiento bayesiano del problema y podemos, designando por Γ_1, Γ_2 las fronteras izquierda y derecha en (4), examinar el criterio sucesivo δ_Γ, $\Gamma = (\Gamma_1, \Gamma_2)$ que se llama *criterio sucesivo de la relación de verosimilitud.* Fue Wald quien lo introdujo por primera vez.

3. Criterio sucesivo que minimiza el número medio de pruebas.

Teorema 2. *Sea* $\Gamma_1 < 1 < \Gamma_2$. *Designemos por* α_1 *y* α_2 *las probabilidades de errores de primero y segundo género del criterio* δ_Γ. *Entonces, entre todos los criterios sucesivos* δ, *para los cuales* $\alpha_1(\delta) \leqslant \alpha_1$, $\alpha_2(\delta) \leqslant \alpha_2$, *el criterio* δ_Γ *tendrá los menores valores* de $\mathbf{M}_1\nu(\delta)$ *y* $\mathbf{M}_2\nu(\delta)$.

Este teorema significa, en particular, que si δ es un criterio construido según la muestra X_n de volumen registrado, para el cual $\alpha_1(\delta) \leqslant \alpha_1$, $\alpha_2(\delta) \leqslant \alpha_2$, entonces

$$\mathbf{M}_i\nu(\delta_\Gamma) \leqslant n, \quad i = 1, 2.$$

Demostración. El criterio bayesiano $\delta_{2,\gamma}$, examinado en el teorema 1, se determina por el conjunto de números (q, a, w_1, w_2). Pero, como ya hemos señalado, la multiplicación de a, w_1, w_2 por un mismo número no altera las fronteras γ_i, así que, de hecho, $\delta_{q,\gamma}$ se determina a base de tres parámetros, por ejemplo, (q, a, w) si se toma $w_1 = w$ y $w_2 = 1 - w$.

Si partimos de este acuerdo, en el teorema 1 hemos construido, a base de los valores dados de (a, w), los números $\gamma_i = \gamma_i(a, w)$ para los cuales el criterio $\delta_{q,\gamma}$ es bayesiano. Ahora necesitaremos, en cierto sentido, la afirmación inversa acerca de que para los valores dados de γ_1, γ_2 existen a, w tales, que $\gamma_i(a, w) = \gamma_i$, o sea, tales a, w, para los cuales el criterio $\delta_{q,\gamma}$ será bayesiano en el problema correspondiente al conjunto (q, a, w). Esta afirmación tiene carácter técnico y se demuestra de un modo bastante complicado (véase [57]). Por eso la aceptamos como tolerable*).

Así pues, examinemos el criterio δ_Γ, y para el valor dado de g hallemos γ_i de las ecuaciones

$$\frac{\gamma_i q}{(1 - \gamma_i)(1 - q)} = \Gamma_i.$$

Para los valores obtenidos de $\gamma_i = \Gamma_i(1 - q)/(\Gamma_i(1 - q) + q)$ hallemos a, w con los cuales el criterio $\delta_{q,\gamma}$ será bayesiano en el problema que corresponde al conjunto (q, a, w). Como $\Gamma_1 < 1 < \Gamma_2$, entonces $\gamma_1 < 1 - q < \gamma_2$ y $\nu(\delta_{q,\gamma}) \geqslant 1$. Esto significa que $\delta_{q,\gamma} = \delta_\Gamma$.

Sea ahora δ cualquier otro criterio para el cual $\alpha_i(\delta) \leqslant \alpha_i$. En vista de que el criterio $\delta_{q,\gamma} = \delta_\Gamma$ minimiza el riesgo bayesiano, entonces

$$q[\alpha_1 w + a\mathbf{M}_1\nu(\delta_\Gamma] + (1 - q)[\alpha_2(1 - w) + a\mathbf{M}_2\nu(\delta_\Gamma)] \leqslant$$
$$\leqslant q[\alpha_1(\delta)w + a\mathbf{M}_1\nu(\delta)] + (1 - q)[\alpha_2(\delta)(1 - w) + a\mathbf{M}_2\nu(\delta)].$$

*) Aquí tampoco demostramos otra afirmación útil acerca de que para las $\mathbf{P}_i$-distribuciones continuas de la magnitud $f_2(X)/f_1(X)$, y para todos los valores dados de α_1, α_2 habrá Γ_1, Γ_2 tales, que $\alpha_1(\delta_\Gamma) = \alpha_1$, $\alpha_2(\delta_\Gamma) = \alpha_2$. Por su esencia esta afirmación se asemeja a los lemas 6.1 y 7.1, pero su demostración es más difícil.

De aquí resulta

$$qM_1\nu(\delta_\Gamma) + (1-q)M_2\nu(\delta_\Gamma) \leqslant qM_1\nu(\delta) + (1-q)M_2\nu(\delta).$$

Como el número $q \in (0, 1)$ aquí es arbitrario, entonces

$$M_1\nu(\delta_\Gamma) \leqslant M_1\nu(\delta), \quad M_2\nu(\delta_\Gamma) \leqslant M_2\nu(\delta). \quad \lhd$$

Aquí hemos utilizado, para la demostración, el mismo método de comparación con los criterios bayesianos que habíamos empleado en los §§ 1, 2, 5.

Examinemos algunas *propiedades del criterio* δ_Γ. Designemos por Ω_i^n los subconjuntos de $\mathscr{X}^\infty$ que se definen del modo siguiente ($X_k = [X_\infty]_k$):

$$\Omega_1^n = \left\{X_\infty:\ \Gamma_1 < \frac{f_2(X_k)}{f_1(X_k)} < \Gamma_2,\ k = 1, \ldots, n-1,\ \frac{f_2(X_n)}{f_1(X_n)} \leqslant \Gamma_1\right\}.$$

El conjunto Ω_2^n se define del mismo modo, pero la última desigualdad debe sustituirse por $f_2(X_n)/f_1(X_n) \geqslant \Gamma_2$. Es evidente que Ω_i^n son disjuntos, pues

$\Omega_i = \bigcup\limits_{n=1}^{\infty} \Omega_i^n$ es la región de aceptación de H_i,

$\nu(\delta_\Gamma) = n$ en la región $\{x \in \mathscr{X}^\infty : x \in \Omega_i^n\}$,

$$\alpha_1(\delta_\Gamma) = \sum_{n=1}^{\infty} \mathbf{P}_1(\Omega_2^n) = \sum_{n=1}^{\infty} \int\limits_{\mathscr{X}^n \cap \Omega_2^n} f_1(x)\mu^n(dx) \leqslant$$

$$\leqslant \sum_{n=1}^{\infty} \int\limits_{\mathscr{X}^n \cap \Omega_2^n} f_2(x)\Gamma_2^{-1}\mu^n(dx) = (1-\alpha_2(\delta_\Gamma))/\Gamma_2. \tag{5}$$

Análogamente se establece que

$$\alpha_2(\delta_\Gamma) \leqslant \Gamma_1(1-\alpha_1(\delta_\Gamma)). \tag{6}$$

Pongamos, para abreviar, $\alpha_i(\delta_\Gamma) = \alpha_i$. El grado de exactitud de las desigualdades obtenidas

$$\Gamma_2 \leqslant \frac{1-\alpha_2}{\alpha_1}, \quad \Gamma_1 \geqslant \frac{\alpha_2}{1-\alpha_1} \tag{7}$$

lo examinaremos más adelante. Ahora aclararemos las propiedades del criterio que obtendremos si hacemos uso de las relaciones (7) en calidad de base para determinar Γ_i por los valores de α_i dados. Si ponemos

$$\Gamma_1' = \frac{\alpha_2}{1-\alpha_1}, \quad \Gamma_2' = \frac{1-\alpha_2}{\alpha_1}, \quad \alpha_i' = \alpha_i(\delta_{\Gamma'}),$$

entonces para el criterio obtenido $\delta_{\Gamma'}$ tendremos, en virtud de (7),

$$\frac{\alpha_2}{1-\alpha_1} \geqslant \frac{\alpha_2'}{1-\alpha_1'}, \quad \frac{1-\alpha_2}{\alpha_1} \leqslant \frac{1-\alpha_2'}{\alpha_1'}, \tag{8}$$

De aquí resulta

$$\alpha_1' \leqslant \frac{\alpha_1(1-\alpha_2')}{1-\alpha_2} \leqslant \frac{\alpha_1}{1-\alpha_2}, \quad \alpha_2' \leqslant \frac{\alpha_2(1-\alpha_1')}{1-\alpha_1} \leqslant \frac{\alpha_2}{1-\alpha_1}.$$

Reduciendo las desigualdades (8) al denominador común y sumándolas, obtenemos asimismo

$$\alpha_1' + \alpha_2' \leqslant \alpha_1 + \alpha_2.$$

Ahora bien, si α_i son pequeños, el criterio $\delta_{\Gamma'}$ tendrá los valores de α_i' cuya suma no excede $\alpha_1 + \alpha_2$, y cada α_i' puede superar α_i sólo insignificantemente y dentro de los límites que conocemos.

Ejemplo 1. Supongamos que x_i tiene una distribución binomial con una probabilidad de éxito p. El problema consiste en verificar la hipótesis $H_1 = \{p = p_1\}$ frente a $H_2 = \{p = p_2\}$, $p_1 < p_2$. En este caso

$$\frac{f_2(X)}{f_1(X)} = \frac{p_2^{\eta_n}(1-p_2)^{n-\eta_n}}{p_1^{\eta_n}(1-p_1)^{n-\eta_n}} = \left(\frac{p_2(1-p_1)}{p_1(1-p_2)}\right)^{\eta_n}\left(\frac{1-p_2}{1-p_1}\right)^n,$$

donde η_n es el número de casos favorables (éxitos) en n pruebas. Para los valores $p_1 = 0{,}05$, $p_2 = 0{,}17$, $\alpha_1 = 0{,}05$, $\alpha_2 = 0{,}10$ obtenemos[*] $\Gamma_1' = 0{,}105$, $\Gamma_2' = 18$, $\alpha_1 = 0{,}031$, $\alpha_2 = 0{,}099$.

$$\mathbf{M}_1\nu(\delta_{\Gamma'}) = 31{,}4, \quad \mathbf{M}_2\nu(\delta_{\Gamma'}) = 30{,}0.$$

Por otro lado, el procedimiento con un volumen fijo de la muestra y con probabilidades de los errores de primero y segundo género correspondientes a 0,05 y 0,10, respectivamente, requiere $n = 57$ observaciones. Ahora bien, en este ejemplo el procedimiento sucesivo reduce casi el doble el número medio de observaciones.

4. Cálculo de los parámetros del mejor criterio sucesivo. Las relaciones (7) y (8) dan la posibilidad de establecer cierta correspondencia entre la frontera Γ y las probabilidades de los errores $\alpha_i(\delta_\Gamma)$. Ahora examinemos más detalladamente el problema de cálculo del criterio δ_Γ.

a) *Fórmulas exactas.* Designemos

$$z_k = \ln\frac{f_2(x_k)}{f_1(x_k)}, \quad k = 1, 2, \ldots$$

$$A_i = \ln\Gamma_i, \quad i = 1, 2.$$

En este caso el criterio δ_Γ puede adquirir la forma siguiente: si $A_1 < 0 < A_2$, entonces los experimentos se realizan sucesivamente, y los valores z_k independientes e igualmente distribuidos se suman hasta que $Z_n = \sum_{k=1}^{n} z_k$ toque por primera vez una de las fronteras A_i. Si es cierta la hipótesis H_2, la divagación descrita será dirigida, por término medio, hacia arriba,

---

[*] Los datos numéricos se han tomado de [57], p. 143.

ya que

$$M_2 z_1 = \int \ln \frac{f_2(x)}{f_1(x)} \cdot f_2(x)\mu(dx) = \varrho_1(P_2 P_1) > 0$$

(véase el lema 2.6.1). De un modo análogo se determina que $M_1 z_1 = -\varrho_1(P_1, P_2) < 0$.

Si las fronteras A_i se alejan a partir del origen de coordenadas, esto corresponde (compárese con (5) y (6)) a la reducción de los errores de primero y segundo género.

Los conjuntos Ω_2^n en los términos de divagación $\{Z_k\}$ tendrán la forma

$$\Omega_2^n = \{A_1 < Z_k < A_2,\ k = 1, \ldots, n-1,\ Z_n \geqslant A_2\}.$$

Los conjuntos Ω_1^n tendrán una forma análoga.

Designemos por $\eta(t)$ la variable aleatoria igual al tiempo de la primera salida de la divagación aleatoria $Z_0 = 0, Z_1, Z_2, \ldots$ fuera de la frontera de t:

$$\eta(t) = \begin{cases} \min\{k: Z_k \geqslant t\} & \text{para } t > 0, \\ \min\{k: Z_k \leqslant t\} & \text{para } t < 0. \end{cases}$$

Es el proceso de reconstrucción que corresponde a la sucesión $\{Z_k\}$ (véase [11], capítulo 8). Las diferencias $\chi(A_i) = Z_{\eta(A_i)} - A_i$ serán los valores de excesos (saltos) a través de los niveles A_i en la divagación $\{Z_k\}$ (véase [11]).

Para la probabilidad de error de primer género ahora podemos escribir

$$\alpha_1(\delta_\Gamma) = \sum_{n=1}^{\infty} P_1(\Omega_2^n) = \sum_{n=1}^{\infty} \int_{\mathcal{X}^n \cap \Omega_2^n} \frac{f_1(x)}{f_2(x)} f_2(x)\mu^n(dx) =$$

$$= \sum_{n=1}^{\infty} M_2(e^{-Z_n}; \Omega_2^n) = \Gamma_2^{-1} M_2(e^{-\chi(A_2)}; \Omega_2), \quad (9)$$

donde $\Omega_2 = \bigcup_{n=1}^{\infty} \Omega_2^n$ es la región de aceptación de H_2. Análogamente

$$\alpha_2(\delta_\Gamma) = \Gamma_1 M_1(e^{\chi(A_1)}; \Omega_1),\ \Omega_1 = \bigcup_{n=1}^{\infty} \Omega_1^n. \quad (10)$$

Seguidamente, para los valores de $M_i\nu$, $i = 1, 2$, $\nu = \nu(\delta_\Gamma)$, en virtud de la identidad de Wald, obtenemos $M_i(Z_\nu) = M_i z_1 M_i \nu$, $i = 1, 2$.

Como $Z_\nu = A_2 + \alpha(A_2)$ en el conjunto Ω_2, $Z_\nu = A_1 + \chi(A_1)$ en el conjunto Ω_1, entonces

$$M_1\nu = \frac{1}{M_1 z_1}[\alpha_1 A_2 + M_1(\chi(A_2); \Omega_2) + (1-\alpha_1)A_1 + M_1(\chi(A_1); \Omega_1)],$$

$$M_2\nu = \frac{1}{M_2 z_1}[(1-\alpha_2)A_2 + M_2(\chi(A_2); \Omega_2) + \alpha_2 A_1 + M_2(\chi(A_1); \Omega_1)]. \quad (11)$$

En varios casos los segundos miembros en las fórmulas (9)—(11) pueden ser determinados de forma explícita. Estas fórmulas también resultan muy útiles en los cálculos aproximados.

b) *Fórmulas (para A_1 y A_2 grandes) y desigualdades aproximadas.* Ya hemos señalado que los grandes valores de $|A_i|$, $i = 1, 2$ corresponden a pequeñas probabilidades de errores $\alpha_i(\delta_\Gamma)$. Examinemos el valor

$$\alpha_1(\delta_\Gamma) = P_1\Big(\sup_{k \leqslant \eta(A_1)} Z_k \geqslant A_2\Big) = P_1\Big(\sup_{k \geqslant 0} Z_k \geqslant A_2\Big) -$$

$$- P_1\Big(\sup_{k \leqslant \eta(A_1)} Z_k < A_2,\ \sup_{k > \eta(A_1)} Z_k \geqslant A_2\Big). \quad (12)$$

Aquí el último sumando no supera, en virtud del carácter markoviano de la variable aleatoria $\eta(t)$, los valores

$$\mathbf{P}_i\Big(\sup_{k>\eta(A_1)} (Z_k - Z_{\eta(A_1)})\Big) \geqslant A_2 - Z_{\eta(A_1)}\Big) \leqslant \mathbf{P}_1\Big(\sup_{k\geqslant 0} Z_k \geqslant A_2 - A_1\Big).$$

Como en casi todos los casos prácticamente interesantes, la probabilidad $u(A) = \mathbf{P}_1(\sup_{k\geqslant 0} Z_k \geqslant A)$ decrece exponencialmente con el aumento de A (véase, por ejemplo, [32], t 2. Esto mismo se puede deducir del capítulo 10 en [11], donde se exponen los métodos de cálculos de $u(A)$*)), entonces, para $|A_i|$ grandes, el valor de $u(A_2 - A_1)$ tendrá un orden más alto de pequeñez que $u(A_2)$. Esto significa, en virtud de (12), que

$$\alpha_1(\delta_T) \approx \mathbf{P}_1\Big(\sup_{k\geqslant 0} Z_k \geqslant A_2\Big) = u(A_2), \tag{13}$$

así que, para grandes A_1 y A_2 en (12), la segunda frontera puede ser omitida. Exactamente igual obtenemos la aproximación

$$\alpha_2(\delta_T) \approx \mathbf{P}_2\Big(\inf_{k\geqslant 0} Z_k \leqslant A_1\Big)). \tag{14}$$

Si $|A_i|$ son grandes y α_i pequeños, los miemboros principales en (11) proporcionan

$$\mathbf{M}_1\nu \approx \frac{A_1}{\mathbf{M}_1 z_1}, \quad \mathbf{M}_2\nu \approx \frac{A_2}{\mathbf{M}_2 z_1}. \tag{15}$$

Estas fórmulas también se basan en la omisión de la segunda frontera (ellas también pueden obtenerse mediante las aproximaciones $\mathbf{M}_i\nu \approx \mathbf{M}_i\eta(A_i) \approx A_i/\mathbf{M}_i z_1$. La última relación tiene lugar en virtud del teorema de reconstrucción ([11])).

Teniendo en cuenta los términos siguientes, según su orden de pequeñez en (11), obtenemos

$$\begin{aligned} \mathbf{M}_1\nu &= \frac{1}{\mathbf{M}_1 z_1}(A_1 + \alpha_1(A_2 - A_1) + \mathbf{M}_1\chi_1), \\ \mathbf{M}_2\nu &\approx \frac{1}{\mathbf{M}_2 z_1}(A_2 + \alpha_2(A_2 - A_1) + \mathbf{M}_2\chi_2), \end{aligned} \tag{16}$$

donde α_i se definen por las aproximaciones (12) y (13), los valores $\mathbf{M}_i\chi_i = \lim_{|A_i|\to\infty} \mathbf{M}_i\chi(A_i)$ pueden ser determinados por los métodos descritos en el capítulo 10 en [11].

Examinemos ahora las desigualdades (8). Como $\chi(A_1) \leqslant 0$, $\chi(A_2) \geqslant 0$, estas igualdades se deducen de (9) y (10) si $\chi(A_i)$ se sustituye por 0. Consiguientemente, la exactitud de tales desigualdades depende del error originado por dicha sustitución.

Si las variables aleatorias z_i están limitadas, $b_1 \leqslant z_1 \leqslant b_2$, es evidente que $\chi(A_2) \leqslant b_2$, $\chi(A_1) \geqslant b_1$, y además de (5) y (6) pueden escribirse las desigualdades inversas. Es decir,

$$\begin{aligned} \alpha_1(\delta_T) &= \Gamma_2^{-1}\mathbf{M}_2(e^{-\chi(A_2)}; \Omega_2) \geqslant \Gamma_2^{-1}e^{-b_2}(1 - \alpha_2), \\ \alpha_2(\delta_T) &\geqslant \Gamma_1 e^{b_1}(1 - \alpha_1). \end{aligned} \tag{17}$$

A fin de ilustrar las relaciones obtenidas, volvamos a examinar el ejemplo 1. Para éste,

$$Z_n = \eta_n \ln\frac{p_2(1 - p_1)}{p_1(1 - p_2)} + n\ln\frac{1 - p_2}{1 - p_1},$$

donde η_n es el número de casos favorables en n pruebas. Esto quiere decir que z_1, para la $\mathbf{P}_i$-distribución, adopta el valor de $b_2 = \ln(p_2/p_1) \approx 1{,}224$ con probabilidad p_i, y el valor de

*) Esto se expone más detalladamente en [9].

$b_i = \ln \frac{1 - p_2}{1 - p_1} \approx -0{,}135$ con probabilidad $1 - p_i$, $i = 1, 2$. De aquí obtenemos

$$M_1 z_1 = -0{,}067, \quad M_2 z_1 = 0{,}096, \quad e^{b_2} = 3{,}400, \quad e^{b_1} = 0{,}874.$$

De los dos últimos valores sólo el segundo es próximo a 1, así que será relativamente exacta tan sólo la segunda igualdad de (17). Utilizando esta desigualdad en (7) para el criterio $\delta_{\Gamma'}$, obtenemos

$$0{,}102 = \frac{\alpha_2'}{1 - \alpha_1'} \leqslant \Gamma_1' \leqslant \frac{\alpha_2'}{(1 - \alpha_1')e^{b_1}} = 0{,}117.$$

Esto proporciona fronteras bastante exactas para el valor de $\Gamma_1' = 0{,}105$. En nuestro caso

$$A_1' = \ln \Gamma_1' \approx -2{,}254, \quad A_2' = \ln \Gamma_2' \approx 2{,}890.$$

De aquí, utilizando las fórmulas aproximadas (15), obtenemos para $M_i\nu'$, $i = 1, 2$, los valores

$$A_1'/M_1 z_1 = 33{,}639, \quad A_2'/M_2 z_1 = 30{,}108.$$

Vemos que incluso aproximaciones que están lejos de ser precisas, tales como (15), dan una noción correcta de las magnitudes $M_i\nu'$. Los resultados serán mucho más exactos si hacemos uso de las fórmulas (16).

§ 12. Verificación de las hipótesis compuestas en el caso general

En este párrafo no vamos a suponer que la muestra pertenece a cualquier familia paramétrica.

El problema de verificación de dos hipótesis en el caso general tiene la forma siguiente. Sean $\mathscr{P}_1$ y $\mathscr{P}_2$ dos familias de distribuciones tales, que la distribución $\mathbf{P}$ de la muestra X pertenece a $\mathscr{P}_1 \cup \mathscr{P}_2$. Se verifica la hipótesis $H_1 = \{X \in \mathbf{P}, \mathbf{P} \in \mathscr{P}_1\}$ frente a $H_2 = \{X \in \mathbf{P}, \mathbf{P} \in \mathscr{P}_2\}$. El principio general de construcción del criterio (no randomizado*)) $\pi(X) = \delta(X)$ aquí queda igual que antes, tal como fue descrito en el § 4 para el caso paramétrico. Se construye precisamente el conjunto crítico $\Omega \subset \mathscr{X}^n$ (que a menudo se identifica con el concepto de criterio) tal, que aceptamos H_2 cuando $X \in \Omega$, y aceptamos H_1 en el caso contrario. El número

$$1 - \varepsilon \inf_{\mathbf{P} \in \mathscr{P}_1} \mathbf{P}(X \notin \Omega)$$

se llama nivel de importancia del criterio. La magnitud

$$\beta_\pi(\mathbf{P}) = \mathbf{P}(X \in \Omega), \quad \mathbf{P} \in \mathscr{P}_2,$$

es el valor de la potencia del criterio π en el "punto" $\mathbf{P} \in \mathscr{P}_2$.

Cuando el conjunto $\mathscr{P}_2$ de alternativas $\mathbf{P}$ es muy abundante, en estas condiciones es muy difícil o incluso imposible comparar las potencias $\beta_\pi(\mathbf{P})$

---

*) Para mantener la uniformidad de las designaciones, en lo sucesivo designaremos los criterios estadísticos con el símbolo π, aunque dentro de los límites de este capítulo se tratará, por lo general, de criterios no randomizados.

de los criterios π y construir los criterios óptimos. Las mínimas exigencias planteadas ante los criterios, en este caso consisten, por lo general, en que para cada $\mathbf{P} \in \mathscr{P}_2$ registrado se cumpla

$$\lim_{n\to\infty} \beta_\pi(\mathbf{P}) = 1.$$

Definición 1. El criterio π que posee esta propiedad se denomina *criterio conciliable.*

La esencia de los criterios sometidos a estudio, al igual que de todos los criterios estadísticos, corresponde al principio fundamental de la estadística matemática, del cual ya hemos hablado en los párrafos 1.4 y 2.31. Si ε es pequeño, entonces, al cumplirse la hipótesis H_1 y al utilizarse muchas veces el criterio construido de nivel $1 - \varepsilon$, nos equivocaremos (o sea, caeremos en la región crítica), por término medio, sólo en el $100\,\varepsilon\%$ de todas las pruebas. Por lo tanto, en caso de cumplirse la hipótesis H_1, consideramos prácticamente imposible la caída en esa región al realizar una sola prueba. Consiguientemente, si a pesar de todo caemos en ella, eso significará que la suposición hecha no es cierta y anunciamos que la hipótesis H_1 no es verdadera. En este caso se dice que los resultados del experimento no concuerdan con la hipótesis H_1 desde el punto de vista del criterio de nivel $1 - \varepsilon$.

Están muy difundidos los criterios de verificación de la hipótesis simple $H_1 = \{X \in \mathbf{P}_1\}$ frente a la hipótesis alternativa compuesta $H_2 = \{X \in \mathbf{P} \neq \mathbf{P}_1\}$; la hipótesis H_2 significa que X es una muestra de la distribución arbitraria $\mathbf{P} \neq \mathbf{P}_1$.

La construcción de los criterios para verificar la hipótesis simple $H_1 = \{X \in \mathbf{P}_1\}$ suele basarse en el "alejamiento" de la distribución empírica $\mathbf{P}_n^*$ respecto a la distribución $\mathbf{P}_1$ desde el punto de vista de cierta "distancia" $d(\mathbf{P}, \mathbf{Q})$. La propiedad deseable de esta distancia consiste en reducir $(\mathbf{P}, Q)$ a cero *sólo* cuando $\mathbf{Q} = \mathbf{P}$, y en transformar la continuidad $d(\mathbf{P}, \mathbf{Q})$ en el "entorno" del punto $\mathbf{Q} = \mathbf{P}$, por ejemplo, en la métrica uniforme (de lo contrario las pequeñas desviaciones de $\mathbf{Q}$ respecto a $\mathbf{P}$ pueden conducir a grandes valores de la distancia d). Recordemos que en el caso paramétrico hemos utilizado consideraciones análogas al construir las estimaciones del parámetro desconocido aplicando el método de distancia mínima.

Así pues, sea $d(\mathbf{P}, \mathbf{Q})$ cierta distancia (no obligatoriamente métrica) en el espacio de distribuciones. Supongamos que a partir de $\varepsilon > 0$ dado se puede hallar tal $c > 0$, para el cual

$$\mathbf{P}_1(d(\mathbf{P}_1, \mathbf{P}_n^*) > c) = \varepsilon.$$

Entonces el criterio se construye del modo siguiente:

$$\pi(X) = \begin{cases} 0, & \text{si } d(\mathbf{P}_1, \mathbf{P}_n^*) \leqslant c, \\ 1, & \text{en el caso contrario.} \end{cases}$$

Evidentemente, π es un criterio de nivel $1 - \varepsilon$.

Al igual que en el § 3, se puede introducir un criterio de *nivel asintótico* $1-\varepsilon$ para el cual

$$\lim_{n\to\infty} \mathbf{P}_1(d(\mathbf{P}_1, \mathbf{P}_n^*) > c) = \varepsilon. \tag{2}$$

Los criterios descritos suelen llamarse *criterios de aceptación* (suponiendo que $\{X \in \mathbf{P}_1\}$). Análogamente, su estructura también puede ser representada de una forma algo diferente. Supongamos que tenemos una funcional $G(\mathbf{P})$ (o una sucesión de funcionales $Gn(\mathbf{P})$) tal, que $G(\mathbf{P}) \neq G(\mathbf{P}_1)$ cuando $\mathbf{P} \neq \mathbf{P}_1$. Entonces podemos poner $\pi(X) = 1$ si $|G(\mathbf{P}_n^*) - G(\mathbf{P}_1)| > c$, y $\pi(X) = 0$ en el caso contrario, donde c se elige partiendo de las mismas consideraciones que en (1) y (2). No es difícil comprobar que este segundo enfoque es equivalente al primero, puesto que a partir de la funcional G se puede construir la distancia $d(\mathbf{P}, \mathbf{P}_1) = |G(\mathbf{P}) - G(\mathbf{P}_1)|$ (compárese con el principio de sustitución en la teoría de estimación), y al contrario, a partir de la distancia $d(\mathbf{P}, \mathbf{P}_1)$ se puede construir la funcional $G(\mathbf{P}) = d(\mathbf{P}, \mathbf{P}_1)\,(G(\mathbf{P}_1) = 0)$ que satisface las propiedades requeridas.

Si en la estructura descrita, la funcional G posee, además, la propiedad $G(\mathbf{P}_n^*) \underset{p}{\to} G(\mathbf{P})$ cuando $X \in \mathbf{P}$ (esto siempre es así cuando G es una función de primero o segundo tipo (véase el § 1.3)), entonces *el criterio construido será conciliable.* En efecto, en este caso el número $c = c(n)$ que asegura la igualdad (2) debe convergir a cero ($\mathbf{P}_1(|G(\mathbf{P}_n^*) - G(\mathbf{P}_1)| > \varepsilon) \to 0$ para cualquier $\varepsilon > 0$) y, por lo tanto, tendremos $G(\mathbf{P}_n^*) \underset{p}{\to} G(\mathbf{P})$, $\mathbf{P}(|G(\mathbf{P}_n^*) - G(\mathbf{P}_1)| > c(n)) \to 0$ para cada $\mathbf{P} \neq \mathbf{P}_1$ registrado.

Examinemos ahora algunos criterios de aceptación bien conocidos que son la realización del enfoque descrito anteriormente.

a) *Criterio de Kolmogórov.* Examinemos la estadística (distancia)

$$D(\mathbf{P}_1, \mathbf{P}_n^*) = \sup_t |F_n^*(t) - F(t)|,$$

donde $F_n^*(t)$ y $F(t)$ son las funciones de distribución que corresponden a las medidas $\mathbf{P}_n^*$ y $\mathbf{P}_1$. En el § 1.8 hemos establecido que si $F(t)$ es continua, $X \in \mathbf{P}_1$, entonces

$$d_K(\mathbf{P}_1, \mathbf{P}_n^*) = \sqrt{n}\, D(\mathbf{P}_1, \mathbf{P}_n^*) \Rightarrow \sup_{0\le t\le 1} |w^\circ(t)|,$$

donde $w^\circ(t)$ es el puente browniano. De aquí se deduce el

Teorema 1 (A.N. Kolmogórov). *Si $F(t)$ es continua, entonces existe*

$$\lim_{n\to\infty} \mathbf{P}_1(d_K(\mathbf{P}_1, \mathbf{P}_n^*) < x) = K(x) \equiv \mathbf{P}\left(\sup_{0\le t\le 1} |w^\circ(t)| < x\right).$$

La función $K(x)$ se puede hallar en forma explícita. La misma es igual a

$$K(x) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2}.$$

Con ayuda de este teorema se pueden construir los criterios de nivel asintótico $1 - \varepsilon$. La función $K(x)$ está tabulada en muchos manuales de estadística matemática. Por eso, para ε dado podemos, mediante tablas, hallar una constante $c = c_\varepsilon$ para la cual $K(c) = 1 - \varepsilon$. Poniendo $\pi(X) = 1$ cuando $d_k(\mathbf{P}_1, \mathbf{P}_n^*) > c_\varepsilon$, obtenemos el criterio de aceptación de nivel asintótico $1 - \varepsilon$. Es fácil notar que el criterio obtenido es conciliable, ya que la funcional $G(\mathbf{P}) = \sup_t |F_P(t) - F(t)|$ (aquí $F_P(t) = \mathbf{P}((-\infty, t))$), con cuya ayuda se ha construido el criterio de Kolmogórov, es continua respecto a F_P en la métrica uniforme y, por consiguiente, es una funcional del tipo II (véase el capítulo para la cual $G(\mathbf{P}_n^*) \underset{c.t.}{\to} G(\mathbf{P})$ cuando $X \in \mathbf{P}$. Queda hacer uso de las observaciones hechas anteriormente sobre las condiciones de conciliabilidad de los criterios de aceptación.

Con ayuda de los resultados del capítulo 1 podemos determinar el comportamiento asintótico de la potencia del criterio de Kolmogórov respecto a alternativas semejantes (véase el § 3). Supongamos que $X \in \mathbf{P}$, donde la distribución $\mathbf{P}$ tiene la función de distribución

$$F_P(x) = F(x) + p(x)n^{-1/2}.$$

Supondremos, para abreviar, que $p(x)$ es continua, y que $F(x)$ es continua y estrictamente monótona. La potencia $\beta(\mathbf{P})$ del criterio de Kolmogórov en el "punto" $\mathbf{P}$ será igual a

$$\beta(\mathbf{P}) = \mathbf{P}(d_k(\mathbf{P}_1, \mathbf{P}_n^*) > c) = \mathbf{P}\left(\sup_t |F(t) - F_n^*(t)|\sqrt{n} > c\right) =$$

$$= \mathbf{P}\left(\sup_t |F_P(t) - p(t)n^{-1/2} - F_n^*|\sqrt{n} > c\right).$$

Si sustituimos $t = F_P^{-1}(u)$, donde F_P^{-1} es una función inversa a F_P, entonces obtenemos la expresión

$$\mathbf{P}\left(\sup_{0 \leq u \leq 1} |u - p(F_P^{-1}(u))n^{-1/2} - F_n^*(F_P^{-1}(u))|\sqrt{n} - c\right). \tag{4}$$

Aquí $U_n^*(u) = F_n^*(F_P^{-1}(u))$ es una función empírica que corresponde a la distribución $U_{0,1}$ uniforme en $[0, 1]$, así que (4) es igual a

$$\mathbf{P}\left(\sup_{0 \leq u \leq 1} |u - U_n^*(u) - p(F_P^{-1}(u))n^{-1/2}|\sqrt{n} - c\right).$$

Además $F_P^{-1}(u) \to F^{-1}(u)$ en virtud de la estricta monotonía de F. De aquí y de la continuidad de p se desprende que

$$\lim_{n \to \infty} \beta(\mathbf{P}) = \mathbf{P}\left(\sup_{0 \leq t \leq 1} |w^\circ(t) - a(t)| > c\right), \text{ donde } a(t) = p(F^{-1}(t)). \tag{5}$$

Se puede mostrar que esta expresión es mínima cuando $a(t) \equiv 0$ ($p \equiv 0$). En este sentido el criterio de Kolmogórov es un criterio no desplazado asintóticamente.

b) *Criterio de Mises—Smirnov* (*criterio* ω^2). Examinemos, en calidad de distancia entre $\mathbf{P}_1$ y $\mathbf{P}_n^*$, la estadística

$$\omega_n^2 = d_{\omega^2}(\mathbf{P}_1, \mathbf{P}_n^*) = n\int (F(x) - F_n^*(x))^2\, dF(x),$$

con cuya ayuda también es posible construir el criterio de aceptación de un nivel dado. En el capítulo 1 hemos demostrado que aquí, al igual que en el caso precedente, es válido el

Teorema 2. *Existe la distribución límite*

$$\lim_{n\to\infty} \mathbf{P}_1(\omega_n^2 < x) = \Omega(x) \equiv \mathbf{P}\left(\int_0^1 (w^\circ(t))^2\, dt < x\right).$$

La función $\Omega(x)$ tiene una forma muy compleja (véase [8]) y aquí no la mostraremos.

Como la funcional

$$G(\mathbf{P}) = \int (F(t) - F_P(t))^2\, dF(t)$$

es una funcional del tipo II (§ 1.3), entonces, conforme a las mismas consideraciones que en el punto a), el criterio ω^2 es conciliable.

Siguiendo los razonamientos del punto anterior, también se puede establecer el comportamiento asintótico de la potencia $\beta(\mathbf{P})$ del criterio ω^2 para las alternativas semejantes de $\mathbf{P}$ de forma (3). De un modo absolutamente análogo obtenemos que

$$\beta(\mathbf{P}) = \mathbf{P}(\omega_n^2 > c) \to \mathbf{P}\left(\int (w^\circ(t) - a(t))^2\, dt > c\right),$$

donde $a(t)$ está definida en (5). El valor límite obtenido es, al igual que en (5), mínimo para $a(t) \equiv 0$, así que el criterio ω^2 también es un criterio no desplazado asintóticamente.

Los dos criterios examinados, al igual que otros criterios de aceptación de la hipótesis $H_1 = \{X \in \mathbf{P}_1\}$, construidos con ayuda de las distancias $d(\mathbf{P}, \mathbf{Q})$, permiten obtener inmediatamente *conjuntos confidenciales* para la función desconocida de distribución $F(x)$ o para la distribución desconocida $\mathbf{P}_1$ de la muestra X. En efecto, la relación (1) (ó (2)) también puede ser interpretada así: la probabilidad de que el c-entorno del "punto" $\mathbf{P}_n^*$ (en sentido de la distancia d) recubra el "punto" $\mathbf{P}_1$ es igual a $1 - \varepsilon$. (Para (2) obtendremos la variante asintótica de esta afirmación). Ello significa (véase el § 8) que el c-entorno del punto $\mathbf{P}_n^*$ no es más que un conjunto confidencial de nivel $1 - \varepsilon$ para la distribución desconocida $\mathbf{P}_1, X \in \mathbf{P}_1$. El criterio de Kolmogórov, por ejemplo, determina tal entorno en términos de las funciones de distribución: el mismo es el conjunto de todas $F(x)$ para las cuales

$$\sup_t |F(t) - F_n^*(t)| \leqslant c_\varepsilon/\sqrt{n},$$

donde c_ε se deduce de (1).

Volvamos a examinar los criterios. Ya hemos señalado que en los niveles asintóticos de significación podemos confiar únicamente cuando son grandes los valores de n. Pero si el volumen de la muestra no es grande, entonces, al construir el criterio (mejor dicho, al determinar $c = c_\varepsilon$) es necesario utilizar las fórmulas exactas para la distribución de $d(\mathbf{P}_1, \mathbf{P}_n^*)$. No obstante, la obtención de tales fórmulas choca, por lo general, con grandes dificultades. En este sentido desempeñan un papel muy importante los llamados criterios *no paramétricos*, basados en estadísticas cuya distribución no de-

pende de la distribución verdadera $\mathbf{P}_1$ (o no depende del parámetro θ cuando $X \in \mathbf{P}_\theta$).

En este caso, las probabilidades $\mathbf{P}_1(d(\mathbf{P}_1, \mathbf{P}_n^*) < x)$ no dependen de $\mathbf{P}_1$ y, por consiguiente, es posible realizarlos una sola vez, hacer las tablas y utilizarlas posteriormente para cualesquiera $\mathbf{P}_1$.

El criterio de Kolmogórov y el criterio ω^2 no son paramétricos. Este hecho fue establecido en el § 1.6.

Los criterios no paramétricos también surgen al verificar dos hipótesis compuestas.

c) *Criterio de signos.* Supongamos que $F(x)$ es la función de distribución para $\mathbf{P}_1$, y que la hipótesis H_1 consiste en que $F(a) = p$ para un punto a dado. Esta es, evidentemente, una hipótesis compuesta. La hipótesis H_2 es suplementaria: $H_2 = \{X \in \mathbf{P}, F_P(a) \neq p\}$. En este caso es natural hacer uso de la estadística siguiente: designemos por $\nu(X)$ el número de observaciones x_i para las cuales el signo de diferencia $x_i - a$ es negativo. En calidad del conjunto crítico Ω examinaremos todas las muestras X para las cuales

$$\nu(X) \notin (c_1, c_2)$$

con ciertos $c_1 < c_2$.

Si la hipótesis H_1 es verdadera, entonces

$$\mathbf{P}_1(\nu(X) = k) = C_n^k p^k (1 - p)^{n-k}.$$

Así pues, para el caso de la hipótesis H_1, la distribución $\nu(X)$ no depende de $\mathbf{P}_1$, ya que nuestro criterio no es paramétrico. Los números c_i han de elegirse de modo que

$$\mathbf{P}(\nu(X) \in (c_1, c_2)) \geqslant 1 - \varepsilon$$

(debido al carácter discreto de $\nu(X)$, aquí puede ser que no se alcance el signo de igualdad). La heterogeneidad en la elección de c_i se puede eliminar exigiendo el no desplazamiento respecto a los cambios de c. En general, este problema es equivalente a la verificación de la hipótesis acerca de que la probabilidad de éxito en el esquema de Bernoulli es igual a p. Análogamente se pueden construir los criterios "unilaterales" para verificar las hipótesis de que $F(a) \leqslant p$.

Si en calidad de generalización del problema examinado verificamos la hipótesis $F(a_i) = p_i$, $i = 1, ..., r$ para los valores dados de a_i y p_i, llegaremos al criterio χ^2 que hemos examinado detalladamente en el § 16.

d) *Criterio de Morán.* Así se llama el siguiente criterio para verificar la hipótesis de que $X \in \mathbf{P}_1$. Sea $x_{(1)}, ..., x_{(n)}$ una serie variacional construida según la muestra X. Supongamos que $\mathbf{P}_1$ tiene una función continua de distribución F, establezcamos la estadística

$$M_n = \sum_{k=0}^{n} [F(x_{(k+1)}) - F(x_{(k)})]^2, \tag{6}$$

donde se adopta $F(x_{(0)}) = 0$, $F(x_{(n+1)}) = 1$. El criterio de Morán rechaza la hipótesis $\{X \in \mathbf{P}_1\}$ si $M_n > c$.

Evidentemente, este parámetro no es paramétrico, ya que $F(x_k) \in U_{0,1}$. Por lo tanto es suficiente examinar el criterio $M_n > c$ basado en la estadística

$$M_n = \sum_{k=0}^{n} (x_{(k+1)} - x_{(k)})^2$$

y destinado a verificar la uniformidad de la distribución de X. En este caso, la utilización de la estadística M_n es natural, ya que la magnitud $\sum_{i=1}^{n} y_i^2$ alcanza su mínimo a condición de que $\sum_{i=1}^{n} y_i = 1$ en el punto $y_1 = \ldots$ $\ldots = y_n = 1/n$.

Para calcular el nivel asintótico del criterio de Morán puede servir la afirmación siguiente:

Teorema 3. *Si* $X \in \mathbf{P}_1$, *entonces*

$$\sqrt{n}(nM_n/2 - 1) \in \Phi_{0,1}.$$

Demostración. Supongamos que $\xi_j \in \Gamma_{\alpha,1}\, j = 1, 2, \ldots$. Entonces $\zeta_k = \sum_{j=1}^{k} \xi_j \in \Pi_{\alpha,k}$ y, en virtud del corolario 1.6.2, la distribución compatible de las diferencias

$$x_{(1)},\ x_{(2)} - x_{(1)},\ \ldots,\ x_{(n)} - x_{(n-1)},\ 1 - x_{(n)}$$

coincide con la distribución compatible

$$\frac{\xi_1}{\zeta_{n+1}}, \frac{\xi_2}{\zeta_{n+1}}, \ldots, \frac{\xi_{n+1}}{\zeta_{n+1}},$$

así que[*)]

$$M_n \underset{d}{=} \zeta_{n+1}^{-2} \sum_{j=1}^{n+1} \xi_j^2.$$

La distribución de M_n no depende de α, y se puede poner $\alpha = 1$. Entonces (véase el § 2.2)

$$\mathbf{M}\xi_j^k = \Gamma(k+1) = k!, \quad \mathbf{D}\xi_j = 1, \quad \mathbf{D}\xi_j^2 = 20,$$

$$\varrho_n \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\xi_j - 1) \in \Phi_{0,1},$$

$$\eta_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\xi_j^2 - 2) \in \Phi_{0,20},$$

---

*) El signo $\underset{d}{=}$ significa la coincidencia de las distribuciones.

Tenemos

$$nM_{n-1} \underset{d}{=} \frac{n\left[2n + \sum_{j=1}^{n}(\xi_j^2 - 2)\right]}{\left[n + \sum_{j=1}^{n}(\xi_j - 1)\right]^2} = \frac{n(2n + \eta_n\sqrt{n})}{(n + \varrho_n\sqrt{n})^2} = \frac{2 + \eta_n n^{-1/2}}{(1 + \varrho_n n^{-1/2})^2},$$

$$(nM_{n-1} - 2)\sqrt{n} = \frac{\eta_n - 4\varrho_n - 2\varrho_n^2 n^{-1/2}}{(1 + \varrho_n n^{-1/2})^2}. \quad (7)$$

Aquí

$$\eta_n - 4\varrho_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\xi_j' + 2), \quad \xi_j' = \xi_j^2 - 4\xi_j,$$

$$\mathbf{M}\xi_j' = -2, \quad \mathbf{D}\xi_j' = \mathbf{M}(\xi_j^4 - 8\xi_j^3 + 16\xi_j^2) - 4 = 4.$$

Por lo tanto, $\eta_n - 4\varrho_n \in \Phi_{0,4}$, así que, en virtud de los teoremas de continuidad de (7), obtenemos

$$\sqrt{n}(nM_{n-1}/2 - 1) \in \Phi_{0,1}.$$

Esto equivale a la afirmación del teorema. ◁.

Citemos ahora las consideraciones que muestran que el criterio de Morán es conciliable. Examinemos la estadística (6) para $X \in \mathbf{P}$, donde $\mathbf{P}$ se distingue de $\mathbf{P}_1$. Una de las distribuciones $\mathbf{P}_1$ o $\mathbf{P}$ puede considerarse, sin limitar la generalidad, uniforme. Supongamos que ella será $\mathbf{P}$. Con respecto a F podemos suponer, para abreviar, que existe una densidad continua $f(t) = F'(t)$ concentrada en [0, 1]. Entonces, para $X \in U_{0,1}$, la parte principal de nM_n será igual a

$$n\sum_{k=0}^{n}[f(x_{(k+1)})(x_{(k+1)} - x_{(k)})]^2 \underset{d}{=} n\sum_{k=1}^{n+1}[f(\zeta_k/\zeta_{n+1})\zeta_k/\zeta_{n+1}]^2. \quad (8)$$

Según la ley fuerte de los grandes números, $k^{-1}\xi_k \underset{c.s.}{\to} 1$ cuando $k \to \infty$. Por eso, a su vez, la parte principal de (8) será igual a

$$\sum_{k=1}^{n} f^2(k/n)\xi_k^2/n.$$

Volviendo a utilizar la ley de los grandes números (o la desigualdad de Chébishev), obtenemos que esta expresión converge, en probabilidad, hacia

$$2\int_0^1 f^2(t)dt \geqslant 2\left(\int_0^1 f(t)dt\right)^2 = 2.$$

Aquí el signo de desigualdad es estricto cuando $f(t) \not\equiv 1$. Esto quiere decir que cuando $X \in \mathbf{P} = \mathbf{U}_{0,1} \neq \mathbf{P}_1$ y cuando $n \to \infty$,

$$\sqrt{n}(nM_n/2 - 1) \underset{p}{\to} \infty,$$

lo cual conduce, en virtud del teorema 3, a la conciliabilidad del criterio de Morán de cualquier nivel registrado $1 - \varepsilon$. ◁

Siendo conciliable, el criterio de Morán *no distingue*, sin embargo, *las hipótesis afines*. Supongamos que $X \in \mathbf{P} = U_{0,1}$,

$$F(t) = t + p(t)n^{-1/2}, \quad t \in [0, 1], \qquad (10)$$
$$p(0) = p(1) = 0,$$

y que la función $p(t)$ es continuamente derivable. Entonces

$$n^{3/2}M_n = n^{3/2}\sum_{k=0}^{n}(x_{(k+1)} - x_{(k)})^2 + 2n\sum_{k=0}^{n}(x_{(k+1)} - x_{(k)}) \times$$
$$\times (p(x_{(k+1)}) - p(x_{(k)})) + \sqrt{n}\sum_{k=0}^{n}p(x_{(k+1)}) - p(x_{(k)}))^2. \qquad (11)$$

La parte principal de la segunda suma aquí es igual a $2n\sum_{k=0}^{n}p'(x_{(k)})(x_{(k+1)}) - x_{(k)})^2$, o bien, en virtud de las mismas consideraciones que en (9),

$$2\sum_{k=1}^{n}p'(k/n)\xi_k^2/n \underset{P}{\to} 4\int_0^1 p'(t)dt = 0.$$

El último sumando en (11) también converge (en probabilidad) a cero, ya que su parte principal coincide (en distribución) con

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}[p'(k/n)]^2\xi_k^2/n,$$

o con $\frac{2}{\sqrt{n}}\int_0^1 [p'(t)]^2dt \to 0$. Lo dicho significa que para la función F en forma de (10), la estadística $n^{3/2}/M_n/2 - \sqrt{n}$ tendrá la misma distribución límite de $\Phi_{0,1}$ que para $F(t) = t$. ◁

Conviene señalar que de este hecho no se deben sacar conclusiones apresuradas de que el criterio de Morán es malo. La cosa consiste en que, sin distinguir las hipótesis afines de forma (10), el criterio de Morán distingue otras hipótesis (que son, en cierto sentido, también afines) las cuales no pueden ser distinguidas por otros criterios examinados en este párrafo. Se trata de las *hipótesis para las densidades*.

Examinemos la hipótesis $H_2 = \{X \in \mathbf{P}\}$, donde la distribución $\mathbf{P}$ tiene una densidad de

$$f(t) = \begin{cases} 2 \text{ cuando } 2k\Delta_n \leqslant t < (2k+1)\Delta_n, \\ 0 \text{ cuando } (2k+1)\Delta_n \leqslant t < (2k+2)\Delta_n, \end{cases} \quad k = 0, 1, \ldots, N-1,$$

donde $\Delta_n = \frac{1}{2N}$, $N = N_n > 0$ es un número entero. Entonces, para $\Delta_n = O(n^{-1/2})$, la función de distribución $F_P(t)$, correspondiente a la distribución $\mathbf{P}$, poseerá la propiedad

$$\sup_t |F_P(t) - t| = O(n^{-1/2}).$$

Esto quiere decir que la hipótesis H_2 *como hipótesis para la función de distribución* será tan próxima a $H_1 = \{X \in U_{0,1}\}$, que los criterios de Kolmogórov y ω^2 no las distinguirán (el valor límite de la potencia en el punto $\mathbf{P}$ coincidirá con el nivel límite del criterio). No obstante, como hipótesis para las *densidades*, las hipótesis H_1 y H_2 se distinguen considerablemente, ya que $\sup |f(t) - 1| = 1$. Como $x_{(0)} = 0$, $x_{(n+1)} = 1$, para $X \in \mathbf{P}$ la estadística M_n superará la magnitud $\Delta_n^2 N = \Delta_n/2$. Por consiguiente, si $n/N = 2n\Delta_n \to \infty$ cuando la

P-probabilidad es igual a 1, tendremos

$$nM_n \to \infty.$$

Fijando el conjunto crítico $\Omega_2 = \{nM_n > 3\}$ obtendremos $\mathbf{P}_1(\Omega_2) \to 0$. Esto significa que cuando $\Delta_n = 0(n^{-1/2})$, $\Delta_n n \to \infty$, el criterio de Morán distinguirá las hipótesis H_1 y H_2 con una probabilidad próxima a 1. Con otras palabras, la estadística M_n es sensible a las desviaciones relacionadas con la densidad, y el propio criterio de Morán puede ser recomendado como criterio para verificar las hipótesis referentes a las *densidades*. Por otro lado, del § 1.10 sabemos que la velocidad con que las densidades empíricas se aproximan a la densidad verdadera es inferior a $n^{-1/2}$. Por eso, la "indistinguibilidad" de las hipótesis de las densidades que difieren una de otra en orden de $n^{-1/2}$ (véase (10)) no debe causar sorpresa.

De acuerdo con el criterio de Morán y con algunos otros criterios examinados anteriormente, se puede hacer una observación general. Si se comparan dos criterios de un mismo nivel registrado, el primero de los cuales está destinado al rechazamiento de mayor número de alternativas que el segundo, la potencia del primer criterio para cada alternativa registrada rechazada por ambos criterios) será, por lo general, menor que la potencia del segundo. A título de ejemplo elemental que ilustra esta circunstancia, el lector puede examinar los criterios $|x_1| > \lambda_{\varepsilon/2}$ y $x_1 > \lambda_\varepsilon$ destinados a verificar, respectivamente, las hipótesis $\{\alpha \neq 0\}$ y $\{\alpha > 0\}$ frente a $\{\alpha = 0\}$, basándose en la observación $x_1 \in \Phi_{\alpha,1}$. Aquí λ_ε es la cuantila de distribución $\Phi_{0,1}$ de orden $1 - \varepsilon$. Las potencias en el punto $\alpha > 0$ serán iguales a

$$1 - \Phi_{0,1}(-\lambda_{\varepsilon+2} - \alpha, \quad \lambda_{\varepsilon+2} - \alpha) < 1 - \Phi(\lambda_\varepsilon - \alpha),$$

respectivamente.

§ 13. Criterios asintóticamente óptimos. Criterio de la relación de verosimilitud como criterio asintóticamente bayesiano para verificar una hipótesis simple frente a otra compuesta

1. Propiedades asintóticas del c.r.v. y del criterio bayesiano. Examinemos el problema de verificación de una hipótesis simple $H_1 = \{X \in \mathbf{P}_{\theta_1}\}$ frente a la hipótesis alternativa $H_2 = \{X \in \mathbf{P}_\theta;\ \theta \neq \theta_1,\ \theta \in \Theta\}$. En los párrafos precedentes hemos visto, en ejemplos, que en este caso el c.u.m.p. no existe, por lo general.

Vamos a examinar el planteamiento "parcialmente bayesiano" del problema que hemos descrito en los §§ 4 y 9. El mismo consiste en la suposición de que θ es escoge en $\Theta_2 = \Theta\{\theta_1\}$ al azar, con una distribución $\mathbf{Q}_2 = \mathbf{Q}$. Se puede considerar que $\mathbf{Q}$ se da en Θ, $\mathbf{Q}(\{\theta_1\}) = 0$. En este caso la distribución de la muestra X se definirá por la densidad "mediada"

$$f_Q(x) = \int f_t(x)\mathbf{Q}(dt). \tag{1}$$

Ahora bien, si se conoce $\mathbf{Q}$, entonces la hipótesis $H_{Q_2} = H_Q$, en virtud de la cual X tiene una distribución de densidad (1), puede considerarse, junto con H_1, como hipótesis simple, y para la construcción del criterio más potente se puede utilizar el lema de Neumann — Pearson.

Resulta que en este caso para "casi todas" las $\mathbf{Q}$ suaves, los criterios más potentes coincidirán asintóticamente con el criterio de la relación de

verosimilitud

$$R(X) = \frac{\sup_{\theta\in\Theta} f_\theta(X)}{f_{\theta_1}(X)} = \frac{f_{\theta^*}(X)}{f_{\theta_1}(X)} > c \tag{2}$$

y, por consiguiente, no dependerán de Q. Este hecho permite considerar como óptimo el criterio hallado al menos en los casos en que se puede suponer que θ en Θ_2 se escoge aleatoriamente, pero desconocemos su distribución Q.

Antes de enunciar el teorema respectivo recordemos algunos resultados que necesitamos y demostremos una afirmación auxiliar. En ella desempeñarán un papel muy importante las propiedades asintóticas conocidas de la relación de verosimilitud. Vamos a examinar inmediatamente el caso del parámetro multidimensional; todo lo necesario para esto se contiene en los §§ 2.28 y 2.29.

Así pues, supongamos que $\theta \in \Theta \subset R^k$, $k \geqslant 1$, y que se cumplen las condiciones de regularidad (RR) cuya enunciación se da en el § 2.28. Supongamos, además, que Q tiene una densidad $q(t)$ respecto a la medida de Lebesgue $\lambda(dt) = dt$.

Según el lema de Neumann — Pearson, el criterio no randomizado más potente $\pi_{Q_1} = \pi_Q$ para verificar H_1 frente a H_Q tendrá la forma siguiente: $\pi_Q(X) = 1$ si

$$X \in \Omega(c) = \left\{x : \frac{f_Q(x)}{f_{\theta_1}(x)} > c\right\}, \quad f_Q(x) = \int q(u)\, f_u(x)du, \tag{3}$$

donde escogeremos $c = c_n$ más tarde, según el nivel dado del criterio.

Los criterios bayesianos para verificar H_1 frente a H_Q también tendrán la misma forma.

Las probabilidades de los errores de primero y segundo género son iguales a

$$\alpha_1(\pi_Q) = \mathbf{P}_{\theta_1}\left(\frac{f_Q(X)}{f_{\theta_1}(X)} > c\right), \quad 1 - \beta(\pi_Q) = \int q(t)\mathbf{P}_t\left(\frac{f_Q(X)}{f_{\theta_1}(X)} \leqslant c\right)dt, \tag{4}$$

respectivamente, donde $\beta(\pi_Q = \int_{\{f_Q(x) \leqslant c f_{\theta_1}(x)\}} f_Q(x)\mu^n(dx)$ es la potencia del criterio más potente.

Podemos escribir las expresiones análogas para el c.r.v. $\hat{\pi}$ que acepta H_Q si se cumple (2):

$$\alpha_1(\hat{\pi}) = \mathbf{P}\theta_1\left(\frac{f_{\theta^*}(X)}{f_{\theta_1}(X)} > c\right),$$

$$\alpha_2(\hat{\pi}) = \int q(t)\mathbf{P}_t\left(\frac{f_{\theta^*}(X)}{f_{\theta_1}(X)} \leqslant c\right)dt = \int_{\{f_{\theta^*}(x) \leqslant c f_{\theta_1}(x)\}} f_Q(x)\mu^n(dx). \tag{5}$$

Pongamos $I = I(\theta_1)$ (el valor de la matriz de información de Fisher en el punto θ_1)

$$\frac{f_Q(X)}{f_{\theta_1}(X)} = \left(\frac{2\pi}{n}\right)^{k/2}\frac{q(\theta_1)}{\sqrt{|I|}}e^{T(X)},$$
$$\frac{f_{\hat\theta}(X)}{f_{\theta_1}(X)} = e^{\hat T(X)}. \tag{6}$$

Entonces las regiones críticas de los criterios π_Q y $\hat\pi$ (véanse (3) y (2)) pueden escribirse, respectivamente, en la forma

$$T(X) > c_Q, \quad \hat T(X) > \hat c. \tag{7}$$

Lema 1. *Supongamos que se cumplen las condiciones* (*RR*) *del* § 2.28, $X \in \mathbf{P}_{\theta_1}$, y que θ_1 *es el punto interior de* Θ. *Entonces*

$$2T(X) = 2\hat T(X)(1 + \varepsilon_n(X)) \Subset H_k, \quad \varepsilon_n(X) \underset{c.s.}{\to} 0.$$

Demostración. La afirmación del lema es el corolario evidente de los teoremas 2.28.4 y 2.28.5. Sólo debemos señalar que $\hat T(X)$ en las designaciones del teorema 2.28.4 no es otra cosa sino $Y(u^*)$ (cuando $\theta = \theta_1$). ◁

2. Carácter asintóticamente bayesiano del c.r.v.

Pasemos a enunciar la afirmación fundamental. Recordemos que cuando estudiamos las propiedades asintóticas de los criterios, en realidad tenemos presente no uno sino toda la sucesión de los criterios $\pi = \pi_n$, donde π_n es el criterio basado en la muestra X_n. Teníamos la misma situación al examinar las propiedades asintóticas de las estimaciones. Ahora bien, aquí y en lo sucesivo, siempre que esto sea necesario, por criterio π entenderemos la sucesión de las funciones $\pi_n(X_n)$ definidas para cada n y $X_n = [X_\infty]_n$.

Definición 1. *El criterio* π *para verificar la hipótesis* $H_1 = \{\theta \in \Theta_1\}$ *frente a* $H_2 = \{\theta \in \Theta_2\}$ *pertenece a la clase* $\tilde K_\varepsilon$ *de los criterios de nivel asintótico* $1 - \varepsilon$ *si*

$$\lim_{n\to\infty} \sup \sup_{\theta\in\Theta_1} \mathbf{M}_\theta \pi(X) \leqslant \varepsilon. \tag{8}$$

En nuestro caso, cuando la hipótesis H_1 es simple y $\Theta_1 = \{\theta_1\}$, la relación (8) se transforma en desigualdad:

$$\lim_{n\to\infty} \sup \mathbf{M}\theta_1 \pi(X) \leqslant \varepsilon.$$

Sea h_ε una cuantila de orden $1 - \varepsilon$ de la distribución χ^2 de k grados de libertad ($\mathbf{H}_k((h_\varepsilon, \infty)) = \varepsilon$). Entonces, del lema 1 se desprende que $\pi_Q \in \tilde K_\varepsilon$, $\hat\pi \in \tilde K_\varepsilon$ si $c_Q = \hat c = h_\varepsilon/2$.

Definición 2. Pongamos $c_Q = h_\varepsilon/2$, de modo que $\pi_Q \in \tilde K_\varepsilon$. El criterio $\pi \in \tilde K_\varepsilon$ se denomina *criterio asintóticamente bayesiano* (c.a.b.) en $\tilde K_\varepsilon$ para verificar la hipótesis $H_1 = \{\theta = \theta_1\}$ frente a H_Q si para las probabilidades

de los errores de segundo género, calculadas para la hipótesis H_Q, es válida la relación

$$\lim_{n\to\infty}\sup \frac{\alpha_2(\pi)}{\alpha_2(\pi_Q)} = \lim_{n\to\infty}\sup \frac{1-\beta(\pi)}{1-\beta(\pi_Q)} = \lim_{n\to\infty}\sup \frac{\mathbf{M}_Q(1-\pi(X))}{\mathbf{M}_Q(1-\pi_Q(X))} = 1.$$

En esta definición hemos utilizado la relación (y no la diferencia) de las probabilidades de los errores de segundo género, ya que $\alpha_2(\pi_Q) \to 0$ cuando $n \to \infty$.

Teorema 1. *Supongamos que se cumplen las condiciones* (*RR*) *y que el punto* θ_1 *es un punto interior de* Θ. *Entonces el criterio de la relación de verosimilitud* $\hat{\pi}$ (véanse (2) y (7)) *para* $\hat{c} = h_\varepsilon/2$ *pertenece a* $\tilde{K}_\varepsilon$ *y es el c.a.b. en* $\tilde{K}_\varepsilon$ *para verificar* H_1 *frente a* H_Q, *cualquiera que sea la distribución* **Q** *cuya densidad* $q(t)$ *es continua y positiva en* Θ. *En este caso*

$$\alpha_2(\hat{\pi}) \sim \alpha_2(\pi_Q \sim \frac{q(\theta_1)}{n^{k/2}\sqrt{|I|}}\, V_k h_\varepsilon^{k/2},$$

donde $I = I(\theta_1)$, V_k *es el volumen de la esfera unitaria en* R^k.

Demostración. Ya hemos demostrado la pertenencia de $\hat{\pi} \in \tilde{K}_\varepsilon$ cuando $\hat{c} = h_\varepsilon/2$. Examinemos ahora los errores de segundo género. En virtud de (4) y (7) tenemos

$$\alpha_2(\pi_Q) = \int\limits_{\{T(x)\leqslant c_Q\}} f_Q(x)\mu^n(dx) = \mathbf{M}_{\theta_1}\left\{\frac{f_Q(X)}{f_{\theta_1}(X)};\ 2T(X) \leqslant h_\varepsilon\right\} =$$
$$= \left(\frac{2\pi}{n}\right)^{k/2}\frac{q(\theta_1)}{\sqrt{|I|}}\mathbf{M}_{\theta_1}\ \{e^{T(X)};\ 2T(X) \leqslant h_\varepsilon\}.$$

Aquí, bajo el signo de esperanza matemática se encuentra la función limitada de $2T$ que es casi por doquier continua respecto a la distribución límite $(\mathbf{H}_k)$. Por eso, cuando $n \to \infty$, $\chi_k^2 \in \mathbf{H}_k$,

$$\mathbf{M}_{\theta_1}(e^{T(X)};\ 2T(X) \leqslant h_\varepsilon) \to \mathbf{M}\left\{e^{\frac{1}{2}\chi^2 k};\ \chi_k^2 \leqslant h_\varepsilon\right\} =$$
$$= (2\pi)^{-k/2}\int\limits_{\{|y|^2\leqslant h_\varepsilon\}} e^{\frac{1}{2}|y|^2 - \frac{1}{2}|y|^2}\,dy_1 \ldots dy_k = (2\pi)^{-k/2} h_\varepsilon^{k/2} V_k.$$

Determinemos ahora el comportamiento asintótico de $\alpha_2(\hat{\pi})$. Designemos $A_n = \{X\colon \pi_Q \neq \hat{\pi}\}$. En virtud del lema 1 $\mathbf{P}_{\theta_1}(A_n) \to 0$. Por eso, del teorema 2.29.5 se deduce que para cualquier N registrado,

$$\sup_{|u|\leqslant N} \mathbf{P}_{\theta+u\sqrt{n}}(A_n) \to 0. \tag{9}$$

Hagamos uso de la representación (véase (5))

$$\alpha_2(\hat{\pi}) = \int q(t)\mathbf{P}_t(\hat{T}(X) \leqslant \hat{c})dt =$$

$$= \int_{|t-\theta_1|\leqslant N/\sqrt{n}} + \int_{|t-\theta_1|>N/\sqrt{n}} \leqslant \int q(t)\mathbf{P}_t(T(X) \leqslant c)dt +$$

$$+ \int_{|t-\theta_1|\leqslant N/\sqrt{n}} q(t)\mathbf{P}_t(A_n)dt + \int_{|t-\theta_1|>N/\sqrt{n}} q(t)\mathbf{P}_t(\hat{T}(X) \leqslant c)dt.$$

En virtud de (9) obtenemos

$$\operatorname*{lím\,sup}_n n^{k/2}\alpha_2(\hat{\pi}) \leqslant \lim_{n\to\infty} n^{k/2}\alpha_2(\pi_Q) +$$

$$+ \max_t q(t)\cdot \operatorname*{lím\,sup}_{n\to\infty} \int_{|t-\theta_1|>N/\sqrt{n}} \mathbf{P}_t\left(\frac{f_{\theta_*}(X)}{f_{\theta_1}(X)} \leqslant c^{\hat{c}}\right)dt.$$

Pero la probabilidad bajo el signo integral no excede

$$\mathbf{P}_t\left(\frac{f_{\theta_1}(X)}{f_t(X)} \geqslant e^{-\hat{c}}\right) \leqslant \exp\{\hat{c}/2 - |t-\theta_1|^2 ng/2\}. \tag{10}$$

Aquí hemos utilizado el teorema 2.28.1. Por consiguiente, la propia integral no excede

$$e^{\hat{c}/2} \int_{|u|>N} e^{-|u|^2 g/2}du \to 0$$

cuando $N \to \infty$. De aquí se deduce que

$$\operatorname*{lím\,sup}_{n\to\infty} n^{k/2}\alpha_2(\hat{\pi}) \leqslant \lim_{n\to\infty} n^{k/2}\alpha_2(\pi_Q). \tag{11}$$

Es evidente que esto equivale a que $\hat{\pi}$ es el c.a.b.

Sólo queda determinar que $\alpha_2(\hat{\pi}) \sim \alpha_2(\pi_Q)$ o que, también en virtud de (11),

$$\operatorname*{lím\,inf}_{n\to\infty} n^{k/2}\alpha_2(\hat{\pi}) \geqslant \lim_{n\to\infty} n^{k/2}\alpha_2(\pi_Q). \tag{12}$$

Para esto, nótese que el criterio π_Q construido es bayesiano y corresponde a la probabilidad a priori q_1 de la hipótesis H_1 que se define por la ecuación (compárense (3) y (6))

$$\frac{q_1}{1-q_1} = \left(\frac{2\pi}{n}\right)^{k/2} \frac{q(\theta_1)}{\sqrt{|I|}}e^{\hat{c}}.$$

Esto quiere decir que la probabilidad del error π_Q se comportará asintóticamente como

$$\varepsilon q_1 + (1-q_1)\alpha_2(\pi_Q) \sim \varepsilon q_1 + \alpha_2(\pi_Q).$$

Si admitimos que (12) no es cierta, obtenemos el criterio $\hat{\pi}$ para el cual la probabilidad del error será menor. Como esto no es posible, (12) queda demostrada. El teorema está demostrado por completo. ◁

De los razonamientos citados se deduce que en las probabilidades de los errores de segundo género hacen el aporte principal los valores aleatorios de θ que entran en el entorno $n^{-1/2}$ del punto θ_1 (con ello se explica el orden de pequeñez $n^{-k/2}$ de estas probabilidades).

Las modificaciones insuficientes de los razonamientos para la demostración del teorema 1 también permiten obtener la afirmación siguiente.

Teorema 2. *Los criterios* π' y π'' con las regiones críticas

$$\begin{aligned}\Omega' &= \{x \in \mathscr{X}^n: n(\hat{\theta}^* - \theta_1)I(\theta_1)(\hat{\theta}^* - \theta_1)^T > h_\varepsilon\},\\ \Omega'' &= \{x \in \mathscr{X}^n: L'(X, \theta_1)I^{-1}(\theta_1)(L'(X, \theta_1))^T > h_\varepsilon\}\end{aligned} \tag{13}$$

son, a la par con $\hat{\pi}$, *los c.a.b. en* $\bar{K}_\varepsilon$. *Esta propiedad se conserva si* $I(\theta_1)$ *en* (13) *se sustituye por* $I(\hat{\theta}^*)$.

Los criterios (13) se obtienen si se utiliza el desarrollo

$$\ln \frac{f_{\hat{\theta}^*}(X)}{f_{\theta_1}(X)} = L(X, \hat{\theta}^*) - L(X, \theta_1)$$

en serie cerca del punto $\hat{\theta}^*$ (véase el teorema 2.28.4). La forma del criterio $\hat{\pi}$ es, en cierto sentido, más cómoda, ya que no está relacionada con la dimensión.

La demostración del teorema 2 se la concedemos al lector.

En el caso unidimensional, el conjunto crítico Ω' (al sustituir $I(\theta_1)$ por $I(\hat{\theta}^*)$) tiene la forma

$$\Omega' = \left\{ |\hat{\theta}^* - \theta_1| > \left[\frac{h_\varepsilon}{n|I(\hat{\theta}^*)|}\right]^{1/2} \right\}, \tag{14}$$

donde, evidentemente, $h_\varepsilon = \lambda_{\varepsilon/2}^2$, $\Phi_{0,1}((-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2})) = 1 - \varepsilon$. Vemos que el criterio π' respectivo (14), que equivale asintóticamente a $\hat{\pi}$, puede interpretarse así: $\pi'(X) = 1$ si θ_1 *no ha caído en el intervalo confidencial de nivel asintótico* $1 - \varepsilon$ *para el parámetro* θ, *construido con ayuda de la e.v.m.* $\hat{\theta}^*$.

Esa misma interpretación también se conservará, evidentemente, en el caso multidimensional; además, los conjuntos confidenciales tendrán forma de elipsoides:

$$(\hat{\theta}^* - \theta)I(\hat{\theta}^*)(\hat{\theta}^* - \theta)^T \leqslant n^{-1}h_\varepsilon.$$

Así pues, vemos que la e.v.m. está estrechamente relacionada con el c.a.b.

Ejemplo 1. Supongamos que $X \in \Pi_\lambda$ y que se verifica la hipótesis $H_1 = \{\lambda = \lambda_1\}$ frente a $H_2\{\lambda \neq \lambda_1\}$. En este caso $\hat{\lambda}^* = \bar{x}^*$, $I(\lambda) = \lambda^{-1}$ y el c.a.b. tendrá la forma

$$(\bar{x} - \lambda_1)^2 > h_\varepsilon\lambda_1/n,$$

donde $\mathbf{H}_1((h_\varepsilon, \infty)) = \varepsilon$.

Ejemplo 2. Supongamos que $X \in \Phi_{\alpha,\sigma^2}$ y que se verifica la hipótesis $H_1 = \{(\alpha, \sigma^2) = (\alpha_1^2, \sigma_1^2)\}$ frente a la alternativa adicional. Aquí $\hat{\alpha}^* = \bar{x}$, $\hat{\sigma}^{2*} = S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, $I(\alpha, \sigma^2) = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & \sigma^{-4/2} \end{pmatrix}$ (véase el § 2.16). Por eso el c.a.b. tiene la forma

$$\frac{(\bar{x} - \alpha_1)^2}{\sigma_1^2} + \frac{(S^2 - \sigma_1^2)^2}{2\sigma_1^4} > \frac{h_\varepsilon}{n},$$

donde $\mathbf{H}_2((h_\varepsilon, \infty)) = \varepsilon$.

3. Carácter de no desplazamiento asintótico del c.r.v. Concluyendo este párrafo estableceremos que el c.r.v. (2) no está asintóticamente desplazado. Recordemos previamente que el criterio π para verificar $H_1 = \{\theta \in \Theta_1\}$ frente a $H_2 = \{\theta \in \Theta_2\}$ se llama criterio no desplazado si

$$\inf_{\theta \in \Theta_2} \mathbf{M}_\theta \pi - \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi \geqslant 0.$$

Definición 3. El criterio π se denomina criterio *asintóticamente no desplazado* si

$$\liminf_{n \to \infty} \left(\inf_{\theta \in \Theta_2} \mathbf{M}_\theta \pi - \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi \right) \geqslant 0.$$

Teorema 3. *El c.r.v.* $\hat{\pi}$ (véase (2), (6) y (7)) *para verificar* $H_1 = \{\theta = \theta_1\}$ *frente a* $H_2 = \{\theta \neq \theta_1\}$ *es un criterio asintóticamente no desplazado.*

Demostración. Como en nuestro caso $\Theta_1 = \{\theta_1\}$ y $\lim_{n\to\infty} \mathbf{M}_{\theta_1} \hat{\pi} = \varepsilon$, es suficiente cerciorarse de que

$$\liminf_{n \to \infty} \inf_{t \in \Theta} \mathbf{M}_t \hat{\pi} = \liminf_{n \to \infty} \inf_{t \in \Theta} \mathbf{P}_t\left(\frac{f_{\theta^*}(X)}{f_{\theta_1}(X)} > e^{\hat{c}}\right) \geqslant \varepsilon, \tag{15}$$

donde $\hat{c} = h_\varepsilon/2$.

De la estimación (10) resulta que existe $N > 0$ tal, que

$$\inf_{|t-\theta_1| \geqslant N/\sqrt{n}} \mathbf{P}_t\left(\frac{f_{\theta^*}(X)}{f_{\theta_1}(X)} > e^{\hat{c}}\right) > \varepsilon.$$

Queda demostrar que $\inf_{|t-\theta_1| \leqslant N/\sqrt{n}} \mathbf{M}_t \hat{\pi} \to \varepsilon$.

Pero, en virtud de los teoremas 2.28.4 y 2.29.3, cuando $X \in \mathbf{P}_t$ uniformemente respecto a u, $|u| \leqslant N$, $u = \sqrt{n}(t - \theta_1)$,

$$\hat{T}(X) \Rightarrow \frac{1}{2}(\xi - u)I(\xi - u)^T, \ \xi \in \Phi_{0,I^{-1}},$$

$$\mathbf{M}_t(\hat{\pi}) \to \mathbf{P}_t\left(\frac{1}{2}(\xi - u)I(\xi - u)^T > \hat{c} = h_\varepsilon/2\right).$$

El segundo miembro aquí alcanza su valor mínimo cuando $u = 0$. Este valor es igual a $\mathbf{P}(\xi/\xi^T > h_\varepsilon) = \varepsilon$. ◁

§ 14. Criterios asintóticamente óptimos para verificar las hipótesis compuestas semejantes

1. Planteamiento del problema y definiciones. En el § 3 hemos estudiado dos enfoques asintóticos del problema de verificación de dos hipótesis simples H_1 y H_2. Si consideramos estas hipótesis fijas, o sea, invariables para el volumen creciente n de la muestra X_n, entonces, al calcular las probabilidades de los errores, llegaremos al problema de las probabilidades de grandes desviaciones, de modo que la probabilidad de uno de los errores, como mínimo, convergerá a cero. De acuerdo con otro enfoque, las hipótesis H_1 y H_2 se consideran como elementos de la sucesión de hipótesis "que se aproximan", en este caso la velocidad de aproximación se escoge de manera que las probabilidades de los errores de primero y segundo género converjan hacia sus propios límites (distintos de 0 y 1). Hemos visto que en el caso paramétrico, los valores del parámetro θ_1 y θ_2, correspondientes a las hipótesis H_1 y H_2, deben distinguirse en orden de $n^{-1/2}$. Cada uno de estos enfoques puede ser justificado conforme a las condiciones concretas.

En el párrafo precedente hemos examinado la distribución $\mathbf{Q}$, no dependiente de n, para el valor alternativo de θ y, como era natural de esperar, hemos obtenido que la probabilidad de un error de segundo género converge a cero como $n^{-k/2}$. Esto se debe al hecho de que a esta probabilidad contribuyen principalmente las hipótesis semejantes para las cuales θ está alejado de θ_1 a una distancia del orden de $n^{-1/2}$ (el volumen de la región que contiene tales θ tendrá precisamente un orden de pequeñez de $n^{-k/2}$).

En este párrafo examinaremos el problema de verificación de las *hipótesis compuestas semejantes*, cuando los valores alternativos del parámetro se aproximan cuando $n \to \infty$. Resulta que en este caso, el problema de verificación de las hipótesis se puede reducir, en cierto sentido, a un problema mucho más simple para la distribución normal.

Pasemos a enunciaciones más exactas. Supongamos que a base de la muestra $X \in \mathbf{P}_\theta$ se comprueba la hipótesis $H_1 = \{\theta \in \Theta_1\}$ frente a $H_2 = \{\theta \in \Theta_2\}$. Fijemos cualquier punto interior θ_1 del conjunto Θ y pongamos

$$\theta = \theta_1 + \gamma n^{-1/2}. \qquad (1)$$

Ahora supongamos que el conjunto Θ_i tiene la forma

$$\Theta_i = \theta_1 + \Gamma_i n^{-1/2}, \qquad (2)$$

donde Γ_i no dependen de n. La notación (2) significa que $\theta \in \Theta_i$ sí y sólo sí en (1) $\gamma \in \Gamma_i$. Las hipótesis $H_i = \{\theta \in \Theta_i\}$ para la condición (1) serán llamadas, al igual que en el § 3, hipótesis *semejantes* (en realidad son una sucesión de hipótesis propias de cada n).

El problema de verificación de las hipótesis semejantes H_l a base de la muestra $X \in \mathbf{P}_\theta$ se llamará *problema A*.

Examinemos ahora otro problema. Sea $Y \in \Phi_{\gamma, I^{-1}}$ una muestra de volumen unitario de la población normal $\Phi_{\gamma, I^{-1}}$ con un vector de valores medios γ y con una matriz de segundos momentos $I^{-1} = I^{-1}(\theta_1)$, donde $I(\theta_1)$ es la matriz de información de Fisher para el problema A en el punto θ_1. Designemos por h_l las hipótesis $\{\gamma \in \Gamma_l\}$. El problema de verificación de las hipótesis h_l a base de una sola observación $Y \in \Phi_{\gamma, I^{-1}}$ se denominará *problema B*.

El hecho extraordinario que permite realizar la reducción antes mencionada consiste, aproximadamente, en lo siguiente. Sea $\pi(Y)$ el criterio óptimo en uno u otro sentido (el c.u.m.p., el criterio bayesiano o el criterio minimax) para verificar h_1 frente a h_2 en el problema B. Y sea $\hat{\theta}^*$, como siempre, la e.v.m. en el problema A, $\gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n}$. Entonces el criterio $\pi(\gamma^*)$ para verificar H_1 frente a H_2 en el problema A poseerá asintóticamente las mismas propiedades que el criterio $\pi(Y)$ en el problema B.

Ahora bien, para hallar el criterio asintóticamente óptimo en el problema A, debemos examinar el problema B, que es más simple, y encontrar en éste (si es posible) el criterio π dotado de la propiedad de optimización necesaria. Si ahora tomamos, en calidad de la observación Y, el valor de γ^* y lo sustituimos en π, obtendremos el criterio buscado en el problema A.

Este hecho podría llamarse *indicio límite de optimización*. Su sentido es bastante sencillo. Pues sabemos, de los resultados del capítulo 2, que cuando $X \in \mathbf{P}_\theta$,

$$\sqrt{n}(\hat{\theta}^* - \theta)I^{1/2}(\theta) \in \Phi_{0,E}$$

uniformemente respecto a θ. Por consiguiente, para $\theta = \theta_1 + \gamma n^{-1/2}$,

$$\sqrt{n}(\hat{\theta}^* - \theta_1) - \gamma \in \Phi_{0, I^{-1}(\theta_1)}$$

o bien, que es lo mismo,

$$\gamma^* \in \Phi_{\gamma, I^{-1}}.$$

Así, pues, $\Phi_{\gamma, I^{-1}}$, o sea, la distribución presente en el problema B no es otra cosa sino la distribución límite para γ^*. Por eso, el indicio límite de optimización es muy natural: reduce el problema de verificación de las hipótesis a un problema "límite". Lo interesante en todo esto es el hecho de que con tal reducción no ocurre ninguna pérdida considerable de información respecto a θ: el criterio óptimo en el problema B también conserva esta optimalidad con arreglo al problema A.

Para conferir a lo dicho un sentido exacto, introduzcamos ahora los principales conceptos de optimización asintótica de los criterios para verificar las hipótesis semejantes en el problema A.

En el párrafo precedente hemos dado la definición de la clase $\tilde{K}_\varepsilon$ de los criterios π de nivel asintótico $1 - \varepsilon$ (definición 2). Para $\pi \in \tilde{K}_\varepsilon$ es válida

$$\lim_{n\to\infty} \sup \sup_{\theta\in\Theta_1} \mathbf{M}_\theta \pi(X) \leqslant \varepsilon.$$

Definición 1. El criterio $\pi_1 \in \tilde{K}_\varepsilon$ se llama *criterio asintóticamente más uniforme y más potente (c.a.u.m.p.) en* $\tilde{K}_\varepsilon$ si para cualquier $\gamma \in \Gamma_2$ y para cualquier $\pi \in \tilde{K}_\varepsilon$

$$\lim_{n\to\infty} \inf (\mathbf{M}_\theta \pi_1(X) - \mathbf{M}_\theta \pi(X)) \geqslant 0,$$

donde $\theta = \theta_1 + \gamma n^{-1/2} \in \Theta_2$ cuando $\gamma \in \Gamma_2$.

Supongamos que en Γ_l se dan las distribuciones $\mathbf{\Pi}_l$ que inducen en Θ_l algunas otras distribuciones (concentradas en el entorno $n^{-1/2}$ del punto θ_1) que designaremos por $\mathbf{Q}_l$, $l = 1, 2$. Las hipótesis de que θ se elige al azar con la distribución $\mathbf{Q}_l$, las designaremos, como antes, por H_{Q_l}.

Por $\tilde{K}_\varepsilon^{Q_1}$ designaremos la clase de criterios π para los cuales

$$\lim_{n\to\infty} \sup \mathbf{M}_{Q_1} \pi(X) \leqslant \varepsilon,$$

donde $\mathbf{M}_{Q_l}$ significa la esperanza matemática incondicional de la distribución compatible de θ y X, $\theta \in \mathbf{Q}_l$, $X \in \mathbf{P}_\theta$. Es evidente que $\tilde{K}_\varepsilon \subset \tilde{K}_\varepsilon{}^Q$ para cualquier $\mathbf{Q}$.

Definición 2. El criterio $\pi_1 \in \tilde{K}_\varepsilon^{Q_1}$ para verificar H_{Q_1} frente a H_{Q_2} se denomina *criterio asintóticamente bayesiano (c.a.b.) en* $\tilde{K}_\varepsilon^{Q_1}$ si para cualquier otro criterio $\pi \in \tilde{K}_\varepsilon^{Q_1}$,

$$\lim_{n\to\infty} \inf (\mathbf{M}_{Q_2} \pi_1(X) - \mathbf{M}_{Q_2} \pi(X)) \geqslant 0. \tag{3}$$

Se puede dar una definición equivalente del carácter bayesiano en la cual en vez de (3) se exige que

$$\lim_{n\to\infty} \inf (\mathbf{M}_{Q_2} \pi_1(X) - \mathbf{M}_{Q_2} \pi_{Q_1Q_2}(X)) \geqslant 0, \tag{4}$$

donde $\pi_{Q_1Q_2}$ es el criterio bayesiano de $\tilde{K}_\varepsilon^{Q_1}$ para verificar las hipótesis H_{Q_1} y H_{Q_2} (o, que es lo mismo, el criterio más potente para verificar H_{Q_1} frente a H_{Q_2} de nivel asintótico $1 - \varepsilon$).

Cabe señalar que la definición 2 se distingue algo de la del c.a.b. que hemos dado en el párrafo anterior (véase la definición 13.2. Allí figura la relación de las probabilidades de los errores, y no su diferencia). Desde el punto de vista de la exposición ulterior, estas definiciones son equivalentes, pero la última de ellas será la más conveniente para nosotros.

Definición 3. El criterio $\pi_1 \in \tilde{K}_\varepsilon$ se llama criterio *asintóticamente minimax* en $\tilde{K}_\varepsilon$ para verificar H_1 frente a H_2 si para cualquier otro criterio $\pi \in \tilde{K}_\varepsilon$ se cumple

$$\lim_{n\to\infty} \inf \left(\inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi_1(X) - \inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi(X) \right) \geqslant 0. \tag{5}$$

Al igual que al examinar los criterios minimax ordinarios (véase el § 9), para evitar consideraciones poco importantes, es cómodo separar los conjuntos Θ_1 y Θ_2 por medio de cierta zona intermedia, de modo que ellos no se toquen. De lo contrario ambos límites inferiores en (5) pueden resultar iguales a ε para cualquier criterio no desplazado asintóticamente π.

De las definiciones citadas se deduce que la propiedad de una u otra optimización asintótica se distingue de la propiedad corriente de esa misma optimización tan sólo por el hecho de que ante la respectiva diferencia aparece el signo lím inf.

A la par con los criterios asintóticamente bayesianos y minimax, en las clases $\bar{K}_\varepsilon$ y $\bar{K}_\varepsilon^{Q_1}$ se puede estudiar las clases asintóticamente bayesianas y minimax ordinarias. Supongamos que en $\Theta = \Theta_1 \cup \Theta_2$ tenemos la distribución $Q = q(1)Q_1 + q(2)Q_2$, $q(1) + q(2) = 1$. Entonces, el criterio π_1 se denomina *asintóticamente bayesiano para la distribución a priori* Q, si para cualquier otro criterio π,

$$\varliminf_{n\to\infty} [q(1)\mathrm{M}_{Q_1}\pi_1(X) + q(2)\mathrm{M}_{Q_2}(1 - \pi_1(X)) - \\ - q(1)\mathrm{M}_{Q_1}\pi(X) - q(2)\mathrm{M}_{Q_2}(1 - \pi(X))] \leqslant 0. \quad (6)$$

La probabilidad de error del criterio π promediado respecto a Q, presente en esta desigualdad, puede ser escrita mediante la probabilidad $\alpha(\pi, \theta)$ de error en el punto θ, en forma de $\mathrm{M}_Q\alpha(\pi, \theta)$, donde

$$\alpha(\pi, \theta) = \begin{cases} \mathrm{M}_\theta\pi(X) & \text{cuando} \quad \theta \in \Theta_1, \\ \mathrm{M}_\theta(1 - \pi(X)) & \text{cuando} \quad \theta \in \Theta_2. \end{cases}$$

Entonces, la desigualdad (6) adopta la forma

$$\varliminf_{n\to\infty} \mathrm{M}_Q[\alpha(\pi_1(X), \theta) - \alpha(\pi(X), \theta)] \leqslant 0.$$

El criterio π_1 será *asintóticamente minimax* si

$$\varliminf_{n\to\infty} [\sup_{\theta\in\Theta} \alpha(\pi_1, \theta) - \sup_{\theta\in\Theta} \alpha(\pi, \theta)] \leqslant 0$$

para cualquier otro criterio π.

El estudio de los criterios asintóticamente bayesianos (en $\bar{K}_\varepsilon^{Q_1}$) y asintóticamente minimax (en $\bar{K}_\varepsilon$), y simplemente el estudio de los criterios asintóticamente bayesianos y minimax es, de hecho, una misma cosa. Por ejemplo, el criterio bayesiano de $\bar{K}_\varepsilon^{Q_1}$ es un criterio bayesiano ordinario para $q(1)$ correspondiente. En este párrafo estudiaremos los criterios de las clases $\bar{K}_\varepsilon$ y $\bar{K}_\varepsilon^{Q_1}$, en tanto que los criterios asintóticamente bayesianos y minimax ordinarios serán examinados en los capítulos ulteriores al investigar un planteamiento más general del problema.

2. Afirmaciones principales. Para simplificar al máximo la exposición posterior, introduciremos una suposición que de ningún modo está relacionada con la esencia de la cuestión y que, si se desea, puede ser retirada,

ya que para ello existen todos los resultados necesarios. Es decir, supondremos que los conjuntos Γ_i están limitados, o sea, existe $N > 0$ tal, que $\Gamma_i \subset \{\gamma: |\gamma| \leqslant N\}$.

Definición 4. Los criterios π_1 y π_2 para verificar las hipótesis semejantes $H_1 = \{\theta \in \Theta_1\}$ y $H_2 = \{\theta \in \Theta_2\}$ a base de la muestra X, se denominan criterios *asintóticamente equivalentes* si

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_1\cup\Theta_2} |\mathbf{M}_\theta \pi_1(X) - \mathbf{M}_\theta \pi_2(X)| = 0. \tag{7}$$

Después de tal suposición podemos poner la región $|\theta - \theta_1| \leqslant N/\sqrt{n}$ bajo el signo sup en (7).

Los criterios asintóticamente equivalentes π_1 y π_2 poseen las propiedades siguientes:

1) *Si* $\pi_1 \in \tilde{K}_\varepsilon$ (o $\tilde{K}_\varepsilon^{Q_1}$), *entonces* $\pi_2 \in \tilde{K}_\varepsilon$ ($\tilde{K}_\varepsilon^{Q_1}$).

2) *Si* π_1 *posee una de las propiedades de la optimización asintótica en las definiciones* 1—3, *el criterio* π_2 *poseerá esa misma propiedad.*

La primera afirmación se deduce de (7) y de la desigualdad

$$\sup_{\theta\in\Theta_1} \mathbf{M}_\theta \pi_2(X) \leqslant \sup_{\theta\in\Theta_1} \mathbf{M}_\theta \pi_1(X) + \sup_{\theta\in\Theta_1} |\mathbf{M}_\theta(\pi_2 - \pi_1)|.$$

La segunda afirmación se demuestra análogamente. Si, por ejemplo, π_1 es asintóticamente minimax, el carácter asintóticamente minimax de π_2 será el corolario de (7) y de la desigualdad

$$\inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi_2(X) \geqslant \inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi_1(X) - \sup_{\theta\in\Theta_2} |\mathbf{M}_\theta(\pi_2 - \pi_1)|. \quad \triangleleft$$

Las condiciones de la equivalencia asintótica de los criterios son establecidas por el

Lema 1. *Supongamos que en el entorno del punto* θ_1 *se cumplen las condiciones* (*RR*), $\pi_i(X) = I_{\{T_n(X) + \varepsilon_{ni}(X) > c\}}$, $i = 1, 2$, *donde para* $X \in \mathbf{P}_\theta$, *tienen lugar las relaciones* $\varepsilon_{ni}(X) \underset{P_{\theta_1}}{\to} 0$, $T_n(X) \Subset \mathbf{G}$, *y la distribución* $\mathbf{G}$ *es continua. Entonces, los criterios* π_1 *y* π_2 *son asintóticamente equivalentes.*

Demostración. $|\mathbf{M}_t \pi_1(X) - \mathbf{M}_t \pi_2(X)| \leqslant \mathbf{P}_t(A_n)$, donde para el suceso $A_n = \{\pi_1(X) \neq \pi_2(X)\}$ se cumple $\mathbf{P}_{\theta_1}(A_n) = \mathbf{P}_{\theta_1}(T_n(X) + \varepsilon_{n1}(X) > c,\ T_n(X) + \varepsilon_{n2}(X) \leqslant c) + \mathbf{P}_n(T_n(X) + \varepsilon_{n1}(X) \leqslant c,\ T_n(X) + \varepsilon_{n2}(X) > c) \to 0$ cuando $n \to \infty$, ya que la distribución límite T_n es continua. Por consiguiente, en virtud del teorema 2.29.5, $\sup_{|t-\theta_1| \leqslant N/\sqrt{n}} \mathbf{P}_t(A_n) \to 0$. $\triangleleft$

El criterio bayesiano de nivel $1 - \varepsilon$ en el problema B para verificar las hipótesis $\mathfrak{H}_{\Pi_i}$ de que γ se elige al azar con la distribución Π_i en Γ_i, $i = 1, 2$, lo designaremos por $\pi_{\Pi_1\Pi_2}(Y)$. Este criterio tiene la forma

$$r(Y) \equiv \frac{\displaystyle\int \exp\left\{-\frac{1}{2}(Y-u)I(Y-u)^T\right\}\Pi_2(du)}{\displaystyle\int \exp\left\{-\frac{1}{2}(Y-u)I(Y-u)^T\right\}\Pi_1(du)} > c, \tag{8}$$

donde $c = c_\varepsilon$ se elige de la condición

$$\int \varphi(\gamma, c)\Pi_1(d\gamma) = \varepsilon, \varphi(\gamma, c) = \mathbf{P}(r(Y) > c), \quad Y \in \Phi_{\gamma, I^{-1}}. \tag{9}$$

Estas relaciones significan, evidentemente, que $\mathbf{M}_{\Pi_1}\pi_{\Pi_1\Pi_2}(Y) = \varepsilon$.

Nótese que $r(y)$ es una función analítica de y. En virtud de su analiticidad, esta función no puede adquirir un valor constante en el conjunto de la medida positiva de Lebesgue o de la medida $\Phi_{\gamma, I^{-1}}$ (de lo contrario sería constante en todas partes, lo cual sólo es posible cuando $\Pi_1 = \Pi_2$). Por lo tanto, $\mathbf{P}(r(Y) = c) = 0$ para cualquier c, y la distribución de $r(Y)$ es continua.

Supongamos, como antes, que $\pi_{Q_1Q_2}(X)$ designa el criterio bayesiano de nivel asintótico $1 - \varepsilon$ en el problema A.

Teorema 1. *Supongamos que las condiciones (RR) se cumplen en el entorno del punto θ_1. Entonces, el criterio $\pi(X) = \pi_{\Pi_1\Pi_2}(\gamma^*)$, $\gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n}$ es asintóticamente equivalente al criterio $\pi_{Q_1Q_2}$ y, por consiguiente, es asintóticamente bayesiano.*

Además,

$$\sup_{|\gamma| \leq N} |\mathbf{M}_{\theta_1 + \gamma/\sqrt{n}}\pi(X) - \varphi(\gamma, c)| \to 0 \tag{10}$$

cuando $n \to \infty$, donde $\varphi(\gamma, c) = \mathbf{M}_\gamma \pi_{\Pi_1\Pi_2}(Y)$ está definida en (9).

Demostración. Examinemos el criterio bayesiano $\pi_{Q_1Q_2}$ en el problema A. Este criterio tiene la forma

$$T(X) = \frac{\int f_{\theta_1 + u/\sqrt{n}}(X)\Pi_2(du)}{\int f_{\theta_1 + u/\sqrt{n}}(X)\Pi_1(du)} > c.$$

Si $X \in \mathbf{P}_{\theta_1}$, entonces, en virtud del teorema 2.28.5,

$$T(X) = r(\gamma^*)(1 + \varepsilon(X, \theta_1))$$

($\gamma^* = u^*$ cuando $\theta = \theta_1$). Como la distribución de $r(Y)$ es continua, $\gamma^* \Rightarrow Y \in \Phi_{0, I^{-1}}$, y como el criterio π tiene la forma $r(\gamma^*) > c$, en virtud del lema 1 queda demostrada la primera afirmación del teorema.

La relación (10) se deduce de la representación

$$\mathbf{M}_{\theta_1 + \gamma/\sqrt{n}}\ \pi(X) = \mathbf{M}_{\theta_1 + \gamma/\sqrt{n}}\ I_{\{r(\gamma^*) > c\}} \to \mathbf{P}(r(Y) > c),$$

$Y \in \Phi_{\gamma, I}$ y del teorema 2.29.4. ◁

Teorema 2. *Supongamos que en el entorno del punto θ_1 se cumplen las condiciones (RR), $\gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n}$.*

Supongamos, además, que existe el criterio minimax $\pi_1(Y)$ de nivel $1 - \varepsilon$ para verificar $\mathfrak{K}_1$ frente a $\mathfrak{K}_2$ en el problema B, y que este criterio es bayesiano

$$\pi_1(Y) = \pi_{\Pi_1\Pi_2}(Y) \tag{11}$$

para las distribuciones a priori Π_1 *y* Π_2 *que satisfacen las condiciones*

$$\mathbf{M}_{\Pi_1}\pi_1(Y) = \sup_{\gamma\in\Gamma_1} \mathbf{M}_\gamma \pi(Y),$$
$$\mathbf{M}_{\Pi_2}\pi_1(Y) = \sup_{\gamma\in\Gamma_2} \mathbf{M}_\gamma \pi(Y), \quad Y\in\Phi_{\gamma,1-\varepsilon} \tag{12}$$

(compárense con las condiciones 9.1). *Entonces, el criterio* $\pi(X) = \pi_{\Pi_1\Pi_2}(\gamma^*)$ *será asintóticamente minimax en la clase* $\tilde{K}_\varepsilon$ *de los criterios para verificar* H_1 *frente a* H_2 *en el problema inicial A.*

Demostración. Como π_1 es un criterio de nivel $1-\varepsilon$, entonces

$$\sup_{\gamma\in\Gamma_1} \mathbf{M}_\gamma \pi_1(Y) = \mathbf{M}_{\Pi_1}\pi(Y) = \varepsilon.$$

De aquí, en virtud de (10) y (12), obtenemos

$$\lim_{n\to\infty} \sup_{\gamma\in\Gamma_1} \mathbf{M}_{\theta_1+\gamma/\sqrt{n}}\, \pi_{Q_1Q_2}(X) = \lim_{n\to\infty} \mathbf{M}_{Q_1}\pi_{Q_1Q_2}(X) = \varepsilon.$$

Esto significa que $\pi_{Q_1Q_2}\in\tilde{K}_\varepsilon$, $\pi_{Q_1Q_2}\in \tilde{K}_\varepsilon^{Q_1}$.

Ahora es necesario demostrar que para cualquier criterio $\pi^*\in\tilde{K}_\varepsilon$,

$$\liminf_{n\to\infty}\left(\inf_{\theta\in\Theta_2}\mathbf{M}_\theta\pi(X) - \inf_{\theta\in\Theta_2}\mathbf{M}_\theta\pi^*(X)\right) \geqslant 0.$$

Tenemos

$$\limsup_{n\to\infty} \inf_{\theta\in\Theta_2} \mathbf{M}_\theta \pi^*(X) \leqslant \limsup_{n\to\infty} \mathbf{M}_\theta\pi^*(X) \leqslant \limsup_{n\to\infty} \mathbf{M}_{Q_2}\pi_{Q_1Q_2}(X). \tag{13}$$

La última desigualdad es válida en virtud del carácter bayesiano de $\pi_{Q_1Q_2}$ (o sea, de la minimización de $q_1\mathbf{M}_{Q_1}\pi_{Q_1Q_2} + (1-q_1)\mathbf{M}_{Q_2}(1-\pi_{Q_1Q_2})$ para q_1 correspondiente) y en virtud del hecho de que $\limsup \mathbf{M}_{Q_1}\pi^*(X)\leqslant\varepsilon$, $\lim \mathbf{M}_{Q_1}\pi_{Q_1Q_2} = \varepsilon$.

Seguidamente, en virtud de (10) y (12) y del teorema 1, el segundo miembro en (13) es igual a

$$\lim_{n\to\infty} \mathbf{M}_{Q_2}\pi_1(\gamma^*) = \mathbf{M}_{\Pi_2}\pi_{\Pi_1\Pi_2}(Y) = \inf_{\gamma\in\Gamma_2}\mathbf{M}_\gamma\pi_{\Pi_1\Pi_2}(Y) =$$
$$= \lim_{n\to\infty} \inf_{\gamma\in\Gamma} \mathbf{M}_{\theta_1+\gamma/\sqrt{n}}\pi_{Q_1Q_2}(X). \quad \lhd$$

Teorema 3. *Supongamos que existe un c.u.m.p.* $\pi_1(Y)$ *de nivel* $1-\varepsilon$ *para verificar* $\mathfrak{k}_1$ *frente a* $\mathfrak{k}_2$ *en el problema B. Supongamos, además, que para cualquier* $\gamma_2\in\Gamma_2$ *existe una distribución* Π_1 *en* Γ_1 *tal, que*

$$\pi_1(Y) = \pi_{\Pi_1\Pi_2}(Y) \tag{14}$$

es el criterio bayesiano para verificar $\mathfrak{k}_{\Pi_1}$ *frente a* $\mathfrak{k}_{\Pi_2}$ (aquí Π_2 está concentrada en el punto γ_2). *Entonces, el criterio* $\pi(X) = \pi_1(\gamma^*)$ *es el c.a.u.m.p. (de nivel asintótico* $1-\varepsilon$*) para verificar* H_1 *frente a* H_2 *en el problema inicial A.*

Nótese que para los problemas de los §§ 5—7 siempre se cumple la condición (14). Esto se deduce de la propia construcción del c.u.m.p. en estos párrafos.

Demostración del teorema 3. La pertenencia de $\pi_1(\gamma^*) \in \tilde{K}_\varepsilon$ se deduce del teorema 1, ya que

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_1} \mathbf{M}_\theta \pi_1(\gamma^*) = \sup_{\theta\in\Theta_1} \lim_{n\to\infty} \mathbf{M}_\theta \pi_1(\gamma^*) = \sup_{\gamma\in\Gamma_1} \varphi(\gamma, c) \leqslant \varepsilon.$$

Sea ahora π^* cualquier otro criterio de $\tilde{K}_\varepsilon$. Entonces

$$\lim_{n\to\infty} \sup \mathbf{M}_{Q_1} \pi^*(X) \leqslant \lim_{n\to\infty} \sup \sup_{\theta\in\Theta_1} \mathbf{M}_\theta \pi^*(X) \leqslant \varepsilon$$

y, por consiguiente, π^* también se puede considerar como criterio de $\tilde{K}_\varepsilon^{Q_1}$ para verificar H_{Q_1} frente a H_{Q_2}, donde Q_1 está inducida por la distribución Π_1 (véase la enunciación del teorema), y Q_2 está concentrada en el punto $\theta_2 = \theta_1 + \gamma_2 n^{-1/2}$. Si $\pi_{Q_1Q_2}$ es un criterio bayesiano de nivel asintótico $1 - \varepsilon$ para estas distribuciones, entonces

$$\lim_{n\to\infty} \mathbf{M}_{\theta_2} \pi_{Q_1Q_2}(X) \geqslant \lim_{n\to\infty} \sup \mathbf{M}_{\theta_2} \pi^*(X).$$

Pero el primer miembro de esta desigualdad coincide, en virtud del teorema 1, con el valor

$$\lim_{n\to\infty} \mathbf{M}_{\theta_2} \pi_{\Pi_1\Pi_2}(\gamma^*) = \lim_{n\to\infty} \mathbf{M}_{\theta_2} \pi_1(\gamma^*). \quad \triangleleft$$

De un modo análogo se puede buscar el c.a.u.m.p. en la clase de los criterios no desplazados asintóticamente.

Observación 1. Si las distribuciones Π_1 y Π_2 están concentradas en los puntos γ_1 y γ_2, respectivamente, entonces

$$r(Y) = \frac{\exp\left\{-\frac{1}{2}(y - \gamma_2)\, I (Y - \gamma_2)^T\right\}}{\exp\left\{-\frac{1}{2}(Y - \gamma_1)\, I (Y - \gamma_1)^T\right\}}.$$

Por lo tanto, la región crítica $\pi_{\Pi_1\Pi_2}(Y)$ tendrá la forma

$$YI(\gamma_2 - \gamma_1)^T = (YI,\ (\gamma_2 - \gamma_1)) > c.$$

En el caso unidimensional, de aquí obtenemos el c.a.m.p. (3.21) que hemos estudiado en el § 3.

Observación 2. Si la distribución Π_1 está concentrada en el punto $u = 0$, y la distribución Π_2 es uniforme en la esfera $|u| \leqslant N$, el denominador de la función $r(Y)$ será igual a $\exp\left\{-\frac{1}{2}YIY^T\right\}$, y el denominador para grandes N y $|\gamma| < N - \sqrt{N}$ será próximo a $\sqrt{|I|}\ (2\pi)^{k/2}$. Por consiguiente, la región crítica para $\pi_{\Pi_1\Pi_2}$ con tales Π_1 y Π_2 será próxima al aspecto exterior del elipsoide

$$YIY^T > c,$$

y la región crítica del criterio asintóticamente bayesiano $\pi_{\Pi_1,\Pi_2}(\gamma^*)$ será próximo a

$$\gamma^* I \gamma^{*T} > c.$$

Esto no es otra cosa sino la forma asintótica del c.r.v. que hemos estudiado en el párrafo anterior (compárese con el teorema 13.2).

Observación 3. En los teoremas 2 y 3 están presentes las condiciones consistentes en que el criterio minimax (teorema 2) o el c.u.m.p. (teorema 3) para el problema B son bayesianos en caso de algunas distribuciones Π_i en Γ_i. En los capítulos posteriores veremos que estas condiciones son inútiles: la clase de todos los criterios bayesianos comprende todos los criterios "inmejorables", incluso los c.u.m.p. y los minimax.

§ 15. Propiedades de la optimización asintótica del criterio de relación de verosimilitud que se deducen del indicio límite de optimización

En este párrafo examinaremos algunas consecuencias de los resultados del § 14, vinculadas con el criterio de relación de verosimilitud. Estableceremos, en particular, la potencia máxima uniforme asintótica y el carácter minimax asintótico del c.r.v. para algunos problemas importantes concretos, relacionados con la verificación de las hipótesis próximas.

En lo sucesivo siempre estimaremos que en el entorno del punto θ_1 se cumplen las condiciones (RR). Para simplificar los cálculos será conveniente, al igual que en el párrafo anterior, considerar, donde sea necesario, que los conjuntos Γ_i están limitados.

1. C.a.u.m.p. para hipótesis semejantes con alternativas unilaterales. Supongamos que el parámetro θ es unidimensional y que se verifica la hipótesis unilateral $H_1 = \{\theta \leqslant \theta_1 + \gamma_1 n^{-1/2}\}$ frente a la hipótesis $H_2 = \{\theta > \theta_2 = \theta_1 + \gamma_2 n^{-1/2}\}$, $\gamma_1 \leqslant \gamma_2$.

Teorema 1. *El criterio de relación de verosimilitud $\hat{\pi}(X)$ con la región crítica*

$$R(X) = \frac{\sup\limits_{\theta\in\Theta_2} f_\theta(X)}{\sup\limits_{\theta\in\Theta_1} f_\theta(X)} > c, \tag{1}$$

cuando $\Theta_1 = \{\theta: \theta \leqslant \theta_1 + \gamma_1 n^{-1/2}\}$, $\Theta_2 = \{\theta: \theta \geqslant \theta_1 + \gamma_2 n^{-1/2}\}$ y con un valor conveniente de c, es asintóticamente equivalente al criterio

$$\gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n} > c_\varepsilon = \lambda_\varepsilon I^{-1/2} + \gamma_1, \quad \Phi_{0,1}(\lambda_\varepsilon) = 1 - \varepsilon \tag{2}$$

y es el c.a.u.m.p. de nivel asintótico $1 - \varepsilon$ para verificar la hipótesis $H_1 = \{\theta \leqslant \theta_1 + \gamma_1 n^{-1/2}\}$ frente a $H_2 = \{\theta > \theta_1 + \gamma_2 n^{-1/2}\}$. En las fórmulas (2), I designa la información de Fisher $I(\theta_1)$ en el punto θ_1 para la familia f_θ.

Demostración. Del § 5 se deduce que para una muestra $Y \in \Phi_{\gamma, I^{-1}}$ de volumen unitario, procedente de una población normal de varianza conocida I^{-1}, existe un c.u.m.p. para verificar la hipótesis $\mathfrak{k}_1 = \{\gamma \leqslant \gamma_1\}$ frente a $\mathfrak{k}_2 = \{\gamma > \gamma_2\}$ de forma $Y > c_\varepsilon$, donde c está definida en (2). Así mismo será, evidentemente, el criterio bayesiano para las distribuciones degeneradas concentradas en los puntos γ_1 y γ_2 (o en los puntos γ_1 y $\gamma > \gamma_1$ si $\gamma_1 = \gamma_2$). A base de esto, del teorema 14.3 se deduce que existe el c.a.u.m.p. de nivel asintótico $1 - \varepsilon$ para verificar H_1 frente a H_2 y que el mismo tiene la forma (2).

Queda demostrar que los criterios (1) y (2) son asintóticamente equivalentes. De acuerdo con el teorema 2.28.4, suponiendo que $Z_1(t) = \dfrac{f_{\theta_1} + t^{(X)}}{f_{\theta_1}(X)}$, tendremos, cuando $X \in \mathbf{P}_{\theta_1}$,

$$R(X) = \frac{\sup\limits_{u > \gamma_2} Z_1(un^{-1/2})}{\sup\limits_{u \leqslant \gamma_1} Z_1(un^{-1/2})} =$$

$$= \frac{\sup\limits_{u>\gamma_2} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(2)}(X)\right\}}{\sup\limits_{u \leqslant \gamma_1} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(1)}(X)\right\}} = T_n(X) + \varepsilon_n^{(3)}(X),$$

donde $\varepsilon_n^{(1)}(X) \xrightarrow[P_{\theta_1}]{} 0, \quad i = 1, 2, 3,$

$$T_n(X) = r(\gamma^*) \equiv \frac{\sup\limits_{u>\gamma_2} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I\right\}}{\sup\limits_{u \leqslant \gamma_1} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I\right\}} =$$

$$= \begin{cases} \exp\left\{-\frac{1}{2}(\gamma^* - \gamma_2)^2 I\right\} & \text{cuando } \gamma^* \leqslant \gamma_1, \\ \exp\left\{-\frac{1}{2}(\gamma^* - \gamma_2)^2 I + \frac{1}{2}(\gamma^* - \gamma_1)^2 I\right\} & \text{cuando } \gamma_1 < \gamma^* < \gamma_2, \\ \exp\left\{\frac{1}{2}(\gamma^* - \gamma_1)^2 I\right\} & \text{cuando } \gamma^* \geqslant \gamma_2. \end{cases}$$

Esta es una función continua monótonamente creciente de γ^*. Por consiguiente, la desigualdad $T_n(X) > c$ equivale a la desigualdad $\gamma^* > c'$ para cierta c'. Además, como $\gamma^* \Rightarrow Y \in \Phi_{0, I^{-1}}$, entonces la distribución $r(Y)$ es absolutamente continua. Las condiciones del lema 14.1 para los criterios (1) y (2) se cumplen. ◁

2. C.a.u.m.p. para alternativas bilaterales. Supongamos que el parámetro θ es, como antes, unidimensional, y que el problema A consiste

en verificar la hipótesis $H_1 = \{(\theta - \theta_1)\sqrt{n} \notin (\gamma_1, \gamma_2)\}$ frente a $H_2 = \{(\theta - \theta_1)\sqrt{n} \in (\gamma_1, \gamma_2)\}$, $\gamma_2 > \gamma_1$. Designemos

$$\bar{\gamma} = \frac{\gamma_1 + \gamma_2}{2}, \quad \Delta = \frac{\gamma_2 - \gamma_1}{2}.$$

Teorema 2. *El criterio de relación de verosimilitud* $\hat{\pi}(X)$, *definido en* (1) *para el valor correspondiente de c y para* $\Theta_1 = \{\theta: (\theta - \theta_1)\sqrt{n} \notin (\gamma_1, \gamma_2)\}$, $\Theta_2 = \{\theta: (\theta - \theta_1)\sqrt{n} \in (\gamma_1, \gamma_2)\}$, *al igual que el criterio*

$$|\gamma^* - \bar{\gamma}| = |(\hat{\theta}^* - \theta_1)\sqrt{n} - \bar{\gamma}| < c_\varepsilon, \tag{3}$$

donde c_ε *se determina de la ecuación* $\Phi_{0,I^{-1}}(-c - \Delta, c - \Delta) = \varepsilon$, *son los c.a.u.m.p. de nivel asintótico* $1 - \varepsilon$ *para verificar* $H_1 = \{(\theta - \theta_1)\sqrt{n} \notin (\gamma_1, \gamma_2)\}$ *frente a* $H_2 = \{(\theta - \theta_1)\sqrt{n} \in (\gamma_1, \gamma_2)\}$.

La demostración de este teorema es bastante parecida a la del teorema anterior. Del § 5 resulta que para el problema *B* destinado a verificar, a base de la observación $Y \in \Phi_{\gamma, I^{-1}}$, la hipótesis $\mathfrak{k}_1 = \{\gamma \notin (\gamma_1, \gamma_2)\}$ frente a $\mathfrak{k}_2 = \{\gamma \in (\gamma_1, \gamma_2)\}$, existe un c.u.m.p. en forma de $c' < Y < c''$, donde c' y c'' se eligen de modo que

$$\Phi_{\gamma_1, I^{-1}}((c', c'')) = \Phi_{\gamma_2, I^{-1}}((c', c'')) = \varepsilon.$$

Es fácil notar que podremos satisfacer estas relaciones si ponemos $c' = \bar{\gamma} - c_\varepsilon$, $c'' = \bar{\gamma} + c_\varepsilon$, ya que

$$\Phi_{\gamma_1, I^{-1}}((\bar{\gamma} - c_\varepsilon, \bar{\gamma} + c_\varepsilon)) = \Phi_{0, I^{-1}}((-c_\varepsilon + \Delta, c_\varepsilon + \Delta)) = \varepsilon,$$
$$\Phi_{\gamma_2, I^{-1}}((\bar{\gamma} - c_\varepsilon, \bar{\gamma} + c_\varepsilon)) = \Phi_{0, I^{-1}}((-c_\varepsilon - \Delta, c_\varepsilon - \Delta)) = \varepsilon.$$

Además, en el § 5 hemos visto que para cualquier $\gamma_0 \in (\gamma_1, \gamma_2)$ existe $q \in (0, 1)$ tal, que el criterio bayesiano π_{Π_1, Π_2} al verificar la hipótesis $\mathfrak{k}_{\Pi_1}$ para la distribución $\Pi_1: \Pi_1(\{\gamma_1\}) = q$, $\Pi_1(\{\gamma_2\}) = 1 - q$ frente a la hipótesis $\mathfrak{k}_{\Pi_2} = \{\gamma = \gamma_0\}$, tendrá la forma

$$c' < Y < c''.$$

Esto significa que las condiciones del teorema 14.3 serán cumplidas y que el criterio (3) será el c.a.u.m.p. para verificar H_1 frente a H_2.

Examinemos ahora el c.r.v. (1) para las regiones Θ_i definidas en el teorema y mostremos que el mismo equivale asintóticamente a (3). Al igual que en la demostración del teorema 1, del teorema 2.28.4 obtenemos que, para $X \in \mathbf{P}_{\theta_1}$,

$$\frac{\sup\limits_{u \in \gamma_1, \gamma_2} Z_1(un^{-1/2})}{\sup\limits_{u \notin \gamma_1, \gamma_2} Z_1(un^{-1/2})} = \frac{\sup\limits_{u \in \gamma_1, \gamma_2} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(1)}(X)\right\}}{\sup\limits_{u \notin \gamma_1, \gamma_2} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(2)}(X)\right\}} = T_n(X) + \varepsilon_n^{(3)}(X),$$

donde $\varepsilon_n^{(i)}(X) \underset{P_{\theta_1}}{\rightarrow} 0$, $i = 1, 2, 3$,

$$T(X) = r(\gamma^*) = \frac{\sup\limits_{u \in \gamma_1, \gamma_2} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I\right\}}{\sup\limits_{u \notin \gamma_1, \gamma_2} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I\right\}} =$$

$$= \begin{cases} \exp\left\{-\frac{1}{2}(\gamma^* - \gamma_1)^2 I\right\} & \text{cuando } \gamma^* \leqslant \gamma_1, \\ \exp\left\{\frac{1}{2}(\gamma^* - \gamma_1)^2 I\right\} & \text{cuando } \gamma_1 < \gamma^* \leqslant \bar{\gamma}, \\ \exp\left\{\frac{1}{2}(\gamma^* - \gamma_2)^2 I\right\} & \text{cuando } \bar{\gamma} < \gamma^* \leqslant \gamma_2, \\ \exp\left\{-\frac{1}{2}(\gamma^* - \gamma_2)^2 I\right\} & \text{cuando } \gamma_2 < \gamma^*. \end{cases}$$

De estas igualdades se deduce que $r(\gamma^*)$ es una función continua monótonamente decreciente de $|\gamma^* - \bar{\gamma}|$ (ella es simétrica respecto al punto $\gamma^* = \bar{\gamma}$). Por eso la desigualdad $r(\gamma^*) > c$ equivale a la desigualdad $|\gamma^* - \bar{\gamma}| < c'$. Como $\gamma^* \Rightarrow Y \in \Phi_{0,I^{-1}}$, entonces se cumplen las condiciones del lema 14.1. ◁

3. **Criterio asintóticamente minimax para hipótesis semejantes referentes a un parámetro multidimensional.** Examinemos ahora el parámetro multidimensional θ. En este caso, el c.a.u.m.p. para verificar la hipótesis $H_1 = \{\theta \in \Theta_1\}$ frente a $H_2 = \{\theta \in \Theta_2\}$, por lo general, no existe, y examinaremos el problema de construcción de los criterios asintóticamente minimax.

Al principio es necesario exponer una observación general para simplificar los razonamientos posteriores. Dicha observación consiste en que el referido problema de verificación de las hipótesis siempre se puede "reparametrizar" (o sea, introducir un nuevo parámetro) de modo que la matriz de información $I = I(\theta_1)$ en el punto θ_1 se convierta en matriz unidad. Para esto es suficiente (véase el § 2.1) efectuar una transformación lineal e introducir un nuevo parámetro β mediante la igualdad

$$\theta = \beta I^{-1/2}.$$

Entonces, la matriz de información de Fisher $J(\beta)$ para la familia paramétrica $\mathbf{P}_{\beta, I^{-1/2}}$ será igual, en el punto $\beta_1 = \theta_1 I^{1/2}$, a

$$J(\beta_1) = I^{-1/2} I I^{-1/2} = E.$$

En este apartado nos será más fácil examinar el parámetro β. Siempre podremos volver al parámetro inicial con ayuda de la transformación lineal inversa.

Así pues, supongamos que $I = I(\theta_1) = E$, y examinemos el problema A de verificación de la hipótesis

$H_1 = \{|\theta - \theta_1| \leqslant an^{-1/2}\}$ frente a $H_2 = \{|\theta - \theta_1| \geqslant bn^{-1/2}\}$, $a < b$ (4)

a base de la muestra $X \in \mathbf{P}_\theta$.

Teorema 3. *El criterio de relación de verosimilitud* $\hat{\pi}$ *definido en* (1) *para el valor correspondiente de* c *y para* $\Theta_1 = \{\theta: |\theta - \theta_1| \leqslant an^{-1/2}\}$ $\Theta_2 = \{\theta: |\theta - \theta_1| \geqslant bn^{-1/2}\}$ *es asintóticamente equivalente, para cualesquiera* $0 \leqslant a < b < \infty$, *a los criterios*

$$\frac{f_{\theta^*}(X)}{f_{\theta_1}(X)} > c, \tag{5}$$

$$|\gamma^*| \equiv |(\hat{\theta}^* - \theta_1)\sqrt{n}| > c_\varepsilon, \tag{6}$$

donde c_ε^2 *es la solución, respecto a* c, *de la ecuación*

$$p_c(a) \equiv \mathbf{P}((\xi_1 + a)^2 + \xi_2^2 + \ldots + \xi_k^2 > c^2) = \varepsilon, \tag{7}$$

y es el criterio asintóticamente minimax de nivel asintótico $1 - \varepsilon$ *para verificar las hipótesis* H_1 *y* H_2 *definidas en* (4). *Las variables aleatorias* ξ_i *en* (7) *son independientes,* $\xi_i \in \Phi_{0,1}$, *la potencia límite garantizada de los criterios* π, (5), (6) *es igual a* $p_{c_\varepsilon}(b)$.

Demostración. Aquí el problema B consistirá en verificar, valiéndose de la observación $Y \in \Phi_{\gamma,E}$, la hipótesis $\hat{H}_1 = \{|\gamma| \leqslant a\}$ frente a $\hat{H}_2 = \{|\gamma| \geqslant b\}$. En el ejemplo 9.1 hemos visto que en este problema existe un criterio minimax de nivel $1 - \varepsilon$ que tiene la forma

$$|Y| > c_\varepsilon.$$

Para construir este criterio hemos utilizado el teorema 9.1. Esto significa que las condiciones del criterio 14.2 se cumplen. Por consiguiente, el criterio

$$|\gamma^*| > c_\varepsilon$$

será un criterio asintóticamente minimax de nivel asintótico $1 - \varepsilon$ para el problema A.

El criterio de relación de verosimilitud (1) aquí tendrá la forma

$$R(X) = \frac{\sup_{|u| \geqslant b} Z_1(un^{-1/2})}{\sup_{|u| \leqslant a} Z_1(un^{-1/2})} > c. \tag{8}$$

Observando exactamente los razonamientos utilizados en las demostraciones de los teoremas 1 y 2, obtendremos que $R(X) = T_n(X) + \varepsilon_n(X)$, $\varepsilon_n(X) \underset{P_{\theta_1}}{\to} 0$, donde

$$T_n(X) = r(\gamma^*) = \frac{\sup\limits_{|u| \geqslant b} \exp\left\{-\frac{1}{2}|\gamma^* - u|^2\right\}}{\sup\limits_{|u| \leqslant a} \exp\left\{-\frac{1}{2}|\gamma^* - u|^2\right\}}.$$

De aquí, como antes, se deduce la continuidad absoluta de la distribución $r(Y)$ y la equivalencia asintótica de los criterios $R(X) > c$ y $T(X) > c$. Este último equivale al criterio

$$|\gamma^*| > c',$$

el cual, cuando $c' = c_\varepsilon$, será un criterio de nivel $1 - \varepsilon$. Según el teorema 14.2 (véase (14.10)), éste tendrá una potencia límite garantizada igual a $p_{c_\varepsilon}(b)$ (véase el teorema 9.2). ◁

Observación 1. Si volvemos al parámetro inicial (hasta la reparametrización que transforma $I(\theta_1)$ en una matriz unidad), obtendremos que la afirmación del teorema será válida respecto a las hipótesis $H_i = \{\theta \in \Theta_i\}$, donde (compárese con el ejemplo 9.2 cuando $\sigma^2 = I^{-1}$)

$$\Theta_1 = \{\theta: (\theta - \theta_1)I(\theta_1)(\theta - \theta_1)^T \leqslant a^2 n^{-1}\},$$
$$\Theta_2 = \{\theta: (\theta - \theta_1)I(\theta_1)(\theta - \theta_1)^T \geqslant b^2 n^{-1}\}.$$

El criterio (6) adoptará la forma

$$(\hat{\theta}^* - \theta_1)I(\theta_1)(\theta^* - \theta_1)^T n > c_\varepsilon^2$$

o bien (véase el teorema 13.2)

$$L'(X, \theta_1)I^{-1}(\theta_1)(L'(X, \theta_1))^T > c_\varepsilon^2. \quad (9)$$

El criterio de relación de verosimilitud no variará, evidentemente, ya que el valor máximo de $f_\theta(X)$ en la región Θ_i no depende de la sustitución de las variables (después de la transformación correspondiente de las regiones de Θ_i).

También cabe señalar que la forma del criterio (9) es, a veces, más cómoda que la del (5) y el (6), puesto que no está relacionada con los cálculos de $\hat{\theta}^*$. Sustituciones análogas pueden hacerse con arreglo a los criterios (2) y (3) en los teoremas 1 y 2. Le dejamos al lector que las haga él mismo.

Observación 2. De un modo absolutamente análogo al teorema 3 se puede construir el criterio asintóticamente minimax para los problemas A que pueden ser reducidos al problema B examinado en el ejemplo 9.5.

Observación 3. En el § 13 hemos construido el criterio asintóticamente bayesiano para verificar la hipótesis $\{\theta = \theta_1\}$ frente a $\{\theta \neq \theta_1\}$, el cual tiene la forma del c.r.v.

$$\frac{f_{\theta^*}(X)}{f_{\theta_1}(X)} > c.$$

Ahora bien, este criterio, siendo el c.a.b., también posee propiedad asintóticamente minimax al verificar la hipótesis $\{\theta = \theta_1\}$ frente a $\{(\theta - \theta_1)I(\theta_1) \times \times (\theta - \theta_1)^T \geqslant b^2 n^{-1}\}$ para cualquier $b > 0$.

4. Criterio asintóticamente minimax de pertenencia de la muestra a una subfamilia paramétrica. Ahora examinaremos el c.r.v. en un problema más complejo de verificación de la hipótesis $H_1 = \{\theta \in \Theta_1\}$ frente a $H_2 = \{\theta \in \Theta_2\}$ cuando la dimensión l del subconjunto Θ_1 es positiva pero me-

nor que $k > 1$. Supongamos que tenemos la función suave $\theta = g(\alpha)$ del parámetro l-dimensional ($l < k$) $\alpha \in A_1 \subset R^l$. La imagen del conjunto A_1 en Θ, engendrada por la aplicación de g, podemos designarla por Θ_1. El problema consiste en verificar la hipótesis $H_1 = \{\theta \in \Theta_1\}$ de que el parámetro θ pertenece a la "curva" Θ_1 (o bien de que $X \in \mathbf{P}_{g(\alpha)}$ para cierto $\alpha \in A_1$) frente a la alternativa adicional $\{X \in \mathbf{P}_\theta; \theta \notin \Theta_1\}$, así que en este caso $\Theta_2 = \Theta \setminus \Theta_1$. Con otras palabras, éste es el problema de verificación de la pertenencia de la muestra X a la subfamilia paramétrica de distribuciones $\{\mathbf{P}_{g(\alpha)}; \alpha \in A_1\}$.

A esta clase de problemas pertenecen, por ejemplo, los problemas ya conocidos de verificación de la hipótesis $\{X \in \Phi_{\alpha_0,\sigma^2}\}$ frente a $\{X \in \Phi_{\alpha,\sigma^2}; \alpha \neq \alpha_0\}$ para un valor de α_0 dado y un valor de σ^2 desconocido, o los problemas de verificación de la hipótesis $\{X \in \Phi_{\alpha,\sigma_0^2}\}$ frente a $\{X \in \Phi_{\alpha,\sigma^2}; \sigma = \sigma_0\}$ para un valor de σ_0 dado y un valor de α desconocido, y otros.

En cuanto a la curva $\theta = g(\alpha)$ en Θ, supondremos que la misma es dos veces continuamente derivable, y que la matriz $G = \|\partial g_i(\alpha)/\partial\alpha_j\|$ ($i = 1, \ldots, k$; $j = 1, \ldots, l$; $g_i(\alpha)$ y α_i son las coordenadas de $g(\alpha)$ y α, respectivamente) tiene el rango l. Esto quiere decir que podemos realizar la sustitución biunívoca derivable del parámetro (la reparametrización del problema) de modo que las primeras l coordenadas (sin limitar la generalidad se puede suponer que las mismas constituyen $\alpha = (\alpha_1, \ldots, \alpha_l)$ determinen la posición del punto θ en la curva Θ_1, y las demás (designémoslas por $\beta = (\beta_1, \ldots, \beta_{k-1})$) que determinen la posición de θ en el "plano" (subespacio), digamos, ortogonal (pero no obligatoriamente) a la "curva" $g(\alpha)$ en el punto α. Entonces, el problema se reduce a la verificación de la hipótesis $\{\beta = 0\}$ frente a $\{\beta \neq 0\}$ siempre que exista el subparámetro "obstaculizador" desconocido α.

En este caso examinaremos las hipótesis semejantes, suponiendo que $\beta = \gamma'' n^{-1/2}$, y comprobaremos la hipótesis $\{\gamma'' = 0\}$ frente $\{\gamma'' \neq 0\}$, o frente a

$$\{\gamma'' M_2(\alpha)\gamma''^T \geqslant b^2\} \tag{10}$$

para $b > 0$ y para cierta matriz definida positivamente $M_2(\alpha)$.

En las coordenadas iniciales, el último problema corresponderá a la verificación de la hipótesis $H_1 = \{\theta \in \Theta_1\}$ frente a las alternativas semejantes, cuando el parámetro θ se sitúe en el entorno $n^{-1/2}$ de la curva Θ_1 y permanezca fuera de cierto "tubo" que contiene Θ_1 y corresponde al conjunto (10). También es posible otra variante de planteamiento del problema de verificación de las hipótesis semejantes, la cual parte del hecho de que el parámetro θ está "localizado" y sabemos que el mismo se halla en el entorno de cierto punto $\theta_0 = g(\alpha^\circ)$, $\alpha^\circ \in A_1$. Entonces, el nuevo parámetro $\tau = (\beta, \alpha - \alpha^\circ)$ será localizado cerca del punto $\tau_0 = (0, 0)$. Pongamos

$\alpha - \alpha^\circ = \gamma' n^{-1/2}$, $\beta = \gamma'' n^{-1/2}$ y comprobemos la hipótesis $\{\gamma'' = 0\}$ frente a $\{\gamma'' \neq 0\}$ o frente a $\{\gamma'' M_2(\alpha^\circ)\gamma''^T \geqslant b^2\}$ al disponer del parámetro localizador γ'.

Los resultados que nos interesan en estos dos planteamientos de los problemas coinciden prácticamente. Sin embargo, es más cómodo investigar el segundo planteamiento, puesto que en este caso disponemos de todos los resultados previos necesarios. La suposición acerca de la localización del parámetro θ tiene carácter convencional, y la forma de las afirmaciones obtenidas más abajo no dependerá de θ_0.

Así pues, consideraremos que el nuevo parámetro $\tau = (\alpha - \alpha^\circ, \beta)$ tiene la forma

$$\tau = \gamma n^{-1/2}, \quad \gamma = (\gamma', \gamma''),$$

y comprobaremos la hipótesis $H_1 = \{\gamma'' = 0\}$ frente a $H_2 = \{\gamma'' M_2 \gamma''^T \geqslant b^2\}$, donde en calidad de $M_2 = M_2(\alpha^\circ)$ tomaremos la matriz de información de Fisher para la familia paramétrica $\{\mathbf{P}_{\theta(0,\beta)}\}$ en el punto $\beta = 0$, donde $\theta(\tau) = \theta\,((\alpha - \alpha^\circ, \beta))$ es la función que reconstruye θ según el valor de $\tau = (\tau', \tau'')$.

Teorema 4. *Supongamos que $\theta_0 = g(\alpha^\circ)$ es un punto interior de θ, y que en el entorno de este punto se cumplen las condiciones (RR). Supongamos también, que la función $g(\alpha)$ es dos veces continuamente derivable en el punto α° y que la matriz $G = \|\partial g_i(\alpha)/\partial\alpha_j\|_{\alpha=\alpha^\circ}$ tiene el rango l. Entonces, para Θ_1 y Θ_2 definidas anteriormente, así como para c correspondiente, el criterio de relación de verosimilitud equivale asintóticamente a los criterios*

$$R_1(X) = \frac{f_{\theta^*}(X)}{f_{g(\hat\alpha^*)}(X)} > e^{h_\varepsilon/2}, \tag{11}$$

$$\begin{aligned} &(\hat\theta^* - g(\hat\alpha^*))I(g(\hat\alpha^*))(\hat\theta^* - g(\hat\alpha^*))^T > h_\varepsilon n^{-1},\\ &(\hat\theta^* - g(\hat\alpha^*))I(\hat\theta^*)(\hat\theta^* - g(\hat\alpha^*))^T > h_\varepsilon n^{-1} \end{aligned} \tag{12}$$

y es el criterio asintóticamente minimax de nivel asintótico $1 - \varepsilon$ para verificar la hipótesis $H_1 = \{\theta \in \Theta_1\} = \{\gamma'' = 0\}$ frente a $H_2 = \{\gamma'' M_2 \gamma''^T \geqslant b^2\}$.

La distribución de la estadística $2 \ln R_1(X)$ para $X \in \mathbf{P}_{g(\alpha^\circ)}$ (o sea, para la hipótesis H_1) converge, cuando $n \to \infty$, hacia la distribución χ^2 de $k - l$ grados de libertad (y, por consiguiente, no depende de f_θ y α°). De acuerdo con esto, h_ε en (11) y (12) significa la cuantila de orden $1 - \varepsilon$ de la distribución $\mathbf{H}_{k-l}$.

La potencia asintótica garantizada del c.r.v. es igual a $\mathbf{P}((\xi_1 + b)^2 + \xi_2^2 + \ldots + \xi_{k-l}^2 > h_\varepsilon)$, donde $\xi_i \in \Phi_{0,1}$ y son independientes.

Vemos que los criterios asintóticamente minimax (11) y (12) no están de ningún modo relacionados con α°.

Observación 4. La hipótesis H_2, en términos del parámetro inicial θ puede ser escrita de la forma siguiente:

$$H_2 = \\ = \{\inf_{\nu'} (\theta - g(\alpha^\circ + \gamma' n^{-1/2}) I(g(\alpha^\circ))(\theta - g(\alpha^\circ + \gamma' n^{-1/2}))^T \geqslant b^2 n^{-1}\}.$$

Recordemos que consideramos limitado el conjunto Γ_l, ya que aquí $(\theta - \theta_0) \leqslant Nn^{-1/2}$, $|\gamma'| \leqslant N$ para cierto $N > 0$.

Observación 5. Como veremos de la demostración, la afirmación del teorema conservará por completo su validez si la hipótesis $H_1 = \{\gamma'' = 0\}$ es sustituida por $H_1 = \{\gamma'' M_2 \gamma''^T \leqslant a^2\}$, $a < b$, con la sustitución respectiva del conjunto Θ_1.

Demostración del teorema 4. En calidad de criterio "principal" aquí examinaremos el criterio (11) equivalente a (1) y más cómodo en cuanto a su forma. Además estableceremos la equivalencia asintótica del mismo respecto al criterio asintóticamente minimax, y luego, su equivalencia asintótica a (12).

Examinemos las distribuciones $\mathbf{P}_\theta$ y $\mathbf{P}_{g(\alpha)}$ como dependientes de los parámetros $\tau = (\tau', \tau'')$ y $\alpha = \tau' + \alpha^\circ$, respectivamente. Pongamos $\tau = \gamma n^{-1/2}$, $\gamma = (\gamma', \gamma'')$, de modo que $\tau' = \gamma' n^{-1/2}$, $\tau'' = \gamma'' n^{-1/2}$, y comprobemos la hipótesis $H_1 = \{\gamma'' = 0\}$ frente a $H_2 = \{\gamma'' M_2 \gamma''^T \geqslant b^2\}$, donde M_2 es la matriz de información de Fisher para la familia $\mathbf{P}_{g(\alpha)}$ en el punto α°. Efectuemos ahora una transformación más del parámetro, semejante a la realizada en el ejemplo 9.4 y la cual convierte las matrices de información en matrices de unidad. Supongamos que $\varrho = \tau\Lambda$ y que, respectivamente, $\delta = \gamma\Lambda$ $(\varrho = \delta n^{-1/2})$, donde Λ es una matriz triangular, semejante a la descrita en el ejemplo 9.4 y la cual posee las propiedades siguientes:

$$J^{-1} = \Lambda^T M^{-1} \Lambda = E, \quad J_2^{-1} = \Lambda_2^T M_2^{-1} \Lambda_2 = E,$$

donde J, M, J_2, M_2 son matrices de información en el punto θ_0 para ϱ, τ, ϱ'', τ'', respectivamente (las tildes superiores y las designaciones tienen el mismo sentido que en τ', τ'', γ', γ''), Λ_2 es la matriz del orden $(k - l) \times (k - l)$, formada por los últimos $k - l$ renglones y columnas de la matriz Λ, de modo que $\varrho'' = \tau'' \Lambda_2$, $\delta'' = \gamma'' \Lambda_2$.

En nuevos parámetros las hipótesis H_1 y H_2 se escribirán de la forma siguiente:

$$H_1 = \{\delta'' = 0\}, \quad H_2 = \{|\delta''| \geqslant b\}.$$

De las propiedades de las transformaciones realizadas se deduce que $\theta = \theta_{(\varrho)}$ es una función biunívoca de ϱ y que todas las familias paramétricas examinadas (incluso con parámetros ϱ', ϱ'') satisfacen las condiciones (RR). Pongamos $\varrho_0 = \theta^{-1}(\theta_0)$ (ésta es la solución de la ecuación $\theta(\varrho) = \theta_0$),

$$Z_0(t) = f_{\theta(\varrho_0 + t)}(X)/f_{\theta_0}(X), \quad Y_0(u) = \ln Z_0(un^{-1/2}).$$

Hagamos uso del teorema 2.29.3. Para $|u| \leqslant \delta_n\sqrt{n}$, $X \in \mathbf{P}_{\theta(\varrho)}$ obtenemos

$$\varrho = \varrho_0 + \delta n^{-1/2},$$

$$Y_0(u) = (\xi_n + \delta, u) - \frac{1}{2}(u, u) + (|u|^2 + |\delta^2|\varepsilon_n(X, u, \delta)), \tag{13}$$

donde $|\varepsilon_n(X, u, \delta)| \leqslant \varepsilon_n(X) \underset{P_{\theta(\varrho)}}{\to} 0$ uniformemente respecto a δ para $|\delta| \leqslant \delta_n\sqrt{n}$, donde δ_n es una sucesión arbitraria que converge a cero. En estas igualdades hemos utilizado el hecho de que la matriz de información para el parámetro ϱ es una matriz unidad. El vector ξ_n es

el vector de las funciones derivadas $n^{-1/2}L(X, \theta(\varrho))$ respecto a ϱ_j en el punto $\varrho = \varrho_0 + \delta n^{-1/2}$, de modo que $\xi_n \in \Phi_{0,E}$ uniformemente respecto a ϱ (respecto a δ) cuando $|\delta| \leqslant \delta_n\sqrt{n}$. (En vista de la suposición de que $(\theta - \theta_0)\sqrt{n}$ está limitada, aquí y más adelante es suficiente establecer la uniformidad de convergencia para $|\delta| \leqslant N$, cuando N se ha registrado arbitrariamente. Sin embargo, nada nos molesta establecer también la uniformidad necesaria en una región más amplia $|\delta| \leqslant \delta_n\sqrt{n} \to \infty$.)

Ahora supongamos que $u = (u', u'')$, $u'' = 0$ en (13). Entonces, según el acuerdo anterior respecto a los símbolos con tildes, podemos escribir

$$Y_0((u', 0)) = (\xi_n' + \delta', u') - \frac{1}{2}(u', u') + (|u'|^2 + |\delta|^2)\varepsilon_n(X, u', \delta). \tag{14}$$

De (13) y (14) se deduce que los valores máximos de $Y_\theta(u)$ y $Y_0(u', 0)$ se alcanzan, respectivamente, para

$$\begin{aligned} u &= (\xi_n + \delta)(E + \varepsilon_n(X, \delta)),\\ u' &= (\xi_n' + \delta')(E + \varepsilon_n^{(1)}(X, \delta)), \end{aligned} \tag{15}$$

donde $\varepsilon_n(X, \delta) \underset{P_{\theta(\varrho)}}{\to} 0$, $\varepsilon_n^{(1)}(X, \delta) \underset{P_{\theta(\varrho)}}{\to} 0$ uniformemente en δ, $|\delta| < \delta_n\sqrt{n}/2$. Tan sólo es necesario notar que la probabilidad de grandes valores de $|\xi_n + \delta|$ es uniformemente pequeña, ya que $\xi_n + \delta \in \Phi_{\delta,E}$ uniformemente en δ, $|\delta| < \delta_n\sqrt{n}$ y $\mathbf{P}_\theta(|\xi_n + \delta| > \delta_n\sqrt{n}) \to 0$ uniformemente en δ, $|\delta| < \delta_n\sqrt{n}/2$.

Volvamos ahora a examinar el c.r.v. Para $\theta = \theta_{(\varrho)}$, $X \in \mathbf{P}_\theta$, $\varrho = \varrho_0 + \delta n^{-1/2}$ tenemos

$$R_1(X) = \frac{\sup_\theta f_\theta(X)}{\sup_\alpha f_{g(\alpha)}(X)} = \frac{\sup_u e^{Y_0(u)}}{\sup_{u'} e^{Y_0((u',0))}} = \frac{\exp\left\{\frac{1}{2}|\xi_n + \delta|^2 + \varepsilon_n(X, \delta)\right\}}{\exp\left\{\frac{1}{2}|\xi_n' + \delta'|^2 + \varepsilon_n^{(1)}(X, \delta)\right\}} = \exp\left\{\frac{1}{2}|\xi_n'' + \delta''|^2 + \varepsilon_n''(X, \delta)\right\}, \tag{16}$$

donde la función ε_n con diferentes índices converge a cero en $\mathbf{P}_\theta$-probabilidad uniformemente cuando $|\delta| < \delta_n\sqrt{n}$:

$$2 \ln R_1(X) \Rightarrow |Y'' + \delta''|^2, \quad Y \in \Phi_{0,E}, \tag{17}$$

uniformemente en δ.

En vista de que para $\theta = g(\alpha)$ con la necesidad de $\delta'' = 0$, de aquí resulta la afirmación del teorema respecto a la estadística $2 \ln R_1(X)$.

Recordemos ahora que (véase el teorema 2.29.3) $\xi_n = u^*(E + \varepsilon_n(X, \delta))$, donde $u^* = (\hat{\varrho}^* - \varrho)\sqrt{n}$, $\hat{\varrho}^*$ es la e.v.m. para el parámetro ϱ. De aquí y de la igualdad $\varrho_0 = 0$, suponiendo $\delta^* = (\hat{\varrho}^* - \varrho_0)\sqrt{n}$, obtenemos

$$\xi_n + \delta = \sqrt{n}(\hat{\varrho}^* - \varrho - \varrho_0) + u^*\varepsilon_n(X, \delta) = \sqrt{n}(\hat{\varrho}^* - \varrho_0) + u^*\varepsilon_n(X, \delta) = \delta^* + u^*\varepsilon_n(X, \delta) \in \Phi_{\delta,E},$$

$$\xi_n'' + \delta'' = (\delta^*)'' + (u^*\varepsilon_n(X, \delta))''.$$

Por lo tanto, el segundo miembro en (16) también puede ser escrito en la forma $\exp\left\{\frac{1}{2}|(\delta^*)''|^2 + \varepsilon_n'''(X, \delta)\right\}$, $\varepsilon_n'''(X, \delta) \underset{P_\theta}{\to} 0$. Esto quiere decir que el criterio

$$|(\delta^*)''|^2 > h_\varepsilon \tag{18}$$

y el c.r.v. son asintóticamente equivalentes, o sea,

$$\lim_{n\to\infty} \sup_\alpha \mathbf{P}_{g(\alpha)}(R_1(X) > e^{h_\varepsilon/2}) = \lim_{n\to\infty} \sup_\alpha \mathbf{P}_{g(\alpha)}(|\delta^*|'' > h_\varepsilon) = \varepsilon,$$

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_2} \mathbf{P}_\theta(R_1(X) > e^{h_\varepsilon/2}) = \lim_{n\to\infty} \sup_{\theta\in\Theta_2} \mathbf{P}_\theta(|\delta^*)|^2 > h_\varepsilon) =$$

$$= \sup_{|\delta''|>b^2} \mathbf{P}(|Y'' + \delta''|^2 > h_\varepsilon) = \mathbf{P}((y_1 + b)^2 + y_2^2 + \ldots + y_{k-l}^2 > h_\varepsilon),$$

donde $y_i \in \Phi_{0,1}$ son independientes.

Demostremos ahora que el criterio (18) es un criterio asintóticamente minimax de nivel asintótico $1 - \varepsilon$. Hagamos uso del teorema 14.2. En nuestro caso, $\delta^* = (\hat{\varrho}^* - \varrho_0)\sqrt{n} \in \Phi_{\delta,E}$. El problema B para $Y \in \Phi_{\delta,E}$ se ha examinado en los ejemplos 9.3 y 9.4. Allí hemos establecido que el criterio

$$|Y''|^2 > h_\varepsilon$$

es minimax y de nivel $1 - \varepsilon$. Por consiguiente, de acuerdo con el teorema 14.2, el criterio (18) es asintóticamente minimax.

Para terminar la demostración nos queda establecer la equivalencia asintótica de (11) y (12). Esta equivalencia se deduce fácilmente de los resultados del § 2.29 y del lema 14.1. ◁

Ejemplo 1. Supongamos que $X \in \Phi_{\lambda,\sigma^2}$, donde λ y σ^2 son parámetros escalares. (Aquí utilizaremos el símbolo λ en vez del α tradicional para que no haya confusión con el argumento de la función $g(\alpha)$). Es necesario verificar la hipótesis $\{\lambda = \lambda_0\}$ frente a $\{\lambda \neq \lambda_0\}$ o frente a $\{|\lambda - \lambda_0| \geqslant bn^{-1/2}\}$, $b > 0$, cuando σ se desconoce. Sabemos que en este caso las e.v.m. tienen la forma siguiente. Si ambas componentes λ y σ^2 del vector $\theta = (\lambda, \sigma^2)$ se desconocen, entonces la e.v.m. para θ es

$$\hat{\theta}^* = (\lambda, \sigma^2)^* = (\bar{x}, S^2), \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Si $\lambda = \lambda_0$, la e.v.m. para σ^2 tiene la forma $(\sigma^2)^* = S_1^2 = \frac{1}{n}\sum(x_i - \lambda_0)^2$, así que $g(\hat{\alpha}^*) = (\lambda_0, S_1^2)$. Como

$$f_\theta(X) = (\sqrt{2\pi\sigma})^{-n}\exp\left\{-(2\sigma^2)^{-1}\sum(x_i - \lambda)^2\right\},$$

el criterio de la relación de verosimilitud (11) tiene la forma

$$S_1^2/S^2 > c.$$

En virtud de la igualdad $S_1^2 = S^2 + (\bar{x} - \lambda_0)^2$, este criterio equivale al criterio

$$|\bar{x} - \lambda_0|/S > c_1. \tag{19}$$

Pero éste es el conocido criterio de Student que hemos examinado anteriormente (las propiedades óptimas de este criterio se exponen en el § 7).

Es fácil comprobar que el criterio (12) tendrá esa misma forma. En efecto, en el § 2.16 hemos visto que la matriz $I(\theta)$ para la familia Φ_{λ,σ^2} tiene la forma

$$I(\theta) = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & (2\sigma^4)^{-1} \end{pmatrix}.$$

En nuestro caso $\hat{\theta}^* - g(\hat{\alpha}^*) = (\bar{x} - \lambda_0,\ S^2 - S_1^2) = (\bar{x} - \lambda_0,\ n(\bar{x} - \lambda_0)^2)$,

$$I^{1/2}(\hat{\theta}^*) = \begin{pmatrix} S^{-1} & 0 \\ 0 & (\sqrt{2}S^2)^{-1} \end{pmatrix}.$$

Como en el primer miembro (12) figura el cuadrado de la norma $|(g(\hat{\alpha}^*) - \hat{\theta}^*)I^{1/2}(\hat{\theta}^*)|^2$, el criterio (12) tendrá la forma

$$\frac{(\bar{x} - \lambda_0)^2}{S^2} + \frac{(\bar{x} - \lambda_0)^4}{2S^4} > c_2$$

que, evidentemente, equivale a (19).

Si en vez de $I(\hat{\theta}^*)$ aquí utilizamos $I(g(\hat{\alpha}^*))$, obtendremos el criterio asintóticamente equivalente

$$|\bar{x} - \lambda_0|/S_1 > c_1.$$

Ejemplo 2. Supongamos que $X \in \Phi_{\lambda,\sigma^2}$. Se necesita verificar la hipótesis $\{\sigma = \sigma_0\}$ frente a $\{|\sigma^2 - \sigma_0^2| \geqslant bn^{-1/2}\}$ cuando se desconoce λ. Aquí, la e.v.m. $\hat{\theta}^*$ para $\theta = (\lambda, \sigma^2)$ será, evidentemente, la misma que en el ejemplo precedente. Si $\sigma = \sigma_0$, entonces $\hat{\lambda}^* = \bar{x}$, de modo que $g(\hat{\alpha}^*) = (\bar{x}, \sigma_0^2)$, $\hat{\theta}^* - g(\hat{\alpha}^*) = (0, \sigma_0^2 - S^2)$.

Los criterios (11) (o, que es lo mismo, el criterio de relación de verosimilitud) tienen la forma

$$(S^2 - \sigma_0^2)^2/\sigma_0^4 > 2h_\varepsilon n^{-1},$$

que equivale, evidentemente, a

$$|S^2/\sigma_0^2 - 1| > \sqrt{2h_\varepsilon n^{-1}},$$

donde $\Phi_{0,1}((h_\varepsilon^{1/2}, \infty)) = \varepsilon/2$. Este criterio también ya fue examinado en el § 7.

§ 16. Criterio χ^2. Verificación de las hipótesis por los datos agrupados

1. Criterio χ^2. Propiedades de optimización asintótica. El criterio χ^2 como tal se destina a verificar, basándose en la muestra X de la distribución polinomial B_θ, $\theta = (\theta_1, \ldots, \theta_r)$, $\sum_{i=1}^{r}\theta_i = 1$, la hipótesis simple $H_1 = \{\theta = p\}$ frente a la alternativa adicional $H_2 = \{\theta \neq p\}$, $p = (p_1, \ldots, p_2)$. La distribución polinomial B_θ se describe por las probabilidades $\theta_i = \mathbf{P}(A_i)$, $i = 1, \ldots, r$, de que se produzca, en cada prueba aislada, uno de los r sucesos disjuntos $A_1, \ldots, A_r$. El elemento x_j de la muestra X de esta distribución puede representarse como uno de los vectores $e_1, \ldots, e_r$ con r coordenadas. La coordenada del vector $e_k(r - 1)$ es igual a cero, y la coordenada del número k es igual a 1. En este caso $x_j = e_k$ si se ha producido el suceso

A_k. Designemos por ν_k el número de veces que se produce el suceso A_k en n pruebas independientes. Entonces $\nu = (\nu_1, ..., \nu_r) = \sum_{i=1}^{n} x_i$ es una estadística suficiente para θ, ya que la función de verosimilitud $f_\theta(X)$ tiene la forma

$$f_\theta(X) = \prod_{i=1}^{r} \theta_i^{\nu_i}. \qquad (1)$$

La estadística χ^2 es, por definición,

$$\chi^2(X) = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i},$$

y el conjunto crítico del criterio χ^2 (la región de aceptación de H_2) tiene la forma

$$\chi^2(X) \geq c,$$

donde c se elige según el nivel de significación establecido.

Ahora examinemos más detalladamente el problema antes enunciado acerca de la verificación de la hipótesis $H_1 = \{\theta = p\}$ frente a $H_2 = \{\theta \neq p\}$.

Está claro que las distribuciones $\{B_\theta\}$ forman una familia paramétrica que no depende del parámetro $k = (r-1)$-dimensional $(\theta_1, ..., \theta_{r-1})$; el valor de θ_r se define por la igualdad $\theta_r = 1 - \sum_{i=1}^{r-1} \theta_i$. El vector $(\theta_1, ..., \theta_{r-1})$, al igual que el $(\theta_1, ..., \theta_r)$, será designado con la letra θ. Esto no provocará equivocaciones. La región θ no es otra cosa sino el simplex $\theta_i \geq 0$, $i = 1, ..., r-1$. $\sum_{i=1}^{r-1} \theta_i \leq 1$. La función logarítmica de verosimilitud $L(X, \theta)$ es igual a

$$L(X, \theta) = \sum_{k=1}^{r} \nu_k \ln \theta_k = \sum_{i=1}^{n} l(x_i, \theta). \qquad (2)$$

La familia $\{B_\theta\}$ satisface las condiciones (A_0), (A_μ), (A_c), y también las condiciones de regularidad (RR) en cualquier punto interior de Θ, o sea, en cualquier punto θ para el cual todos $\theta_i > 0$. Efectivamente, en nuestro caso

$$l(x_1, \theta) = \ln \theta_j \quad \text{para } x_1 = e_j;$$

$$\frac{\partial l(x_1, \theta)}{\partial \theta_j} = \begin{cases} \theta_j^{-1} & \text{para } x_1 = e_j, \\ -\theta_r^{-1} & \text{para } x_1 = e_r, \\ 0 & \text{para } x_1 \neq e_j,\ x_1 \neq e_r, \end{cases} \qquad (3)$$

$$\frac{\partial^2 l(x_1, \theta)}{\partial\theta_i\partial\theta_j} = \begin{cases} \dfrac{\delta_{ij}}{\theta_i\theta_j} & \text{para } x_1 = e_j, \\ -\theta_r^{-2} & \text{para } x_1 = e_r, \\ 0 & \text{para } x_1 \neq e_j,\ x_1 \neq e_r, \end{cases} \tag{4}$$

donde δ_{ij} es el símbolo de Kronecker. De estas fórmulas se deduce que

$$\frac{\partial^2 l(x_1, \theta)}{\partial\theta_i\partial\theta_j} = -\frac{\partial l(x_1, \theta)}{\partial\theta_i}\cdot\frac{\partial l(x_1, \theta)}{\partial\theta_j},\quad i, j \leqslant r - 1.$$

Parte de las condiciones (RR) relacionadas con la existencia de las esperanzas matemáticas, aquí se cumplen evidentemente, ya que en nuestro caso el conjunto χ es finito.

De (3) o (4) se deduce

$$I(\theta) = \| I_{ij}(\theta) \| = -\left\| \mathbf{M}_\theta \frac{\partial^2 l(x_1, \theta)}{\partial\theta_i\partial\theta_j} \right\| = \left\| \frac{\delta_{ij}}{\theta_i} + \frac{1}{\theta_r} \right\|, \tag{5}$$
$$i, j = 1, \dots, r - 1.$$

Si en esta matriz sustraemos la primera fila de todas las demás y luego utilizamos el desarrollo en elementos de la primera fila, obtenemos

$$|I(\theta)| = \left(1 + \sum_{j=1}^{r-1} \frac{\theta_j}{\theta_r}\right) \prod_{j=1}^{r-1} \theta_j^{-1} = \left(\prod_{j=1}^{r} \theta_j\right)^{-1}.$$

Así pues, $0 < |I(\theta)| < \infty$ si $\prod_{k=1}^{r} \theta_k > 0$, o sea, si el punto θ es el punto interior del simplex Θ.

Por lo tanto, vemos que podemos utilizar los resultados de los §§ 13 y 14 en los criterios asintóticamente óptimos. De estos resultados se desprende que para verificar la hipótesis $H_1 = \{\theta = p\}$ frente a $H_2 = \{\theta \neq p\}$ existe un c.a.b. que coincide con el criterio de relación de verosimilitud

$$\frac{f_{\theta^*}(X)}{f_p(X)} > c. \tag{6}$$

Este mismo criterio será asintóticamente minimax para verificar H_1 frente a la hipótesis $\{(\theta - p)I(\theta)(\theta - p)^T > b^2 n^{-1}\}$ (véase el teorema 15.3).

Para hallar de una forma más cómoda la región crítica (6), es necesario calcular el valor de $f_{\theta^*}(X)$. Derivando (2) respecto a $\theta_1, \dots, \theta_{r-1}$, obtenemos

$$\frac{\partial L(X, \theta)}{\partial\theta_i} = \frac{\nu_i}{\theta_i} - \frac{\nu_r}{\theta_r},\quad i = 1, \dots, r - 1.$$

Igualando a cero estas derivadas, obtenemos que la e.v.m. equivale a

$$\hat{\theta}^* = n^{-1}\nu,$$

Así que $\hat{\theta}_i^* = n^{-1}\nu_i$.

Ahora bien, pasando a los logaritmos, el criterio (6) se puede escribir de la forma siguiente:

$$\psi^2(X) = \sum_{i=1}^{r} \nu_i \ln \frac{\nu_i}{np_i} > c_1. \tag{7}$$

De acuerdo con el teorema 13.1 (véase también el lema 13.1), la estadística $2\psi^2(X)$ para la hipótesis H_1 tiene una distribución límite χ^2 con $r - 1$ grados de libertad. Por eso obtendremos el criterio de nivel asintótico $1 - \varepsilon$ si ponemos $c_1 = h_\varepsilon/2$, donde h_ε es la cuantila de la distribución H_{r-1} del orden de $1 - \varepsilon$.

¿Qué representa en nuestras condiciones el criterio π' asintóticamente equivalente a (6), obtenido en el teorema 13.2 y que tiene la forma

$$n(\hat{\theta}^* - p)I(p)(\hat{\theta}^* - p)^T > h_\varepsilon? \tag{8}$$

Para $t = (t_1, ..., t_{r-1})$, $s = \sum_{i=1}^{r-1} t_i$, obtenemos

$$tI(p) = \left(\frac{t_1}{p_1} + \frac{s}{p_r}, ..., \frac{t_{r-1}}{p_{r-1}} + \frac{s}{p_r}\right),$$

$$tI(p)t^T = \sum_{i=1}^{r-1} \frac{t_i^2}{p_i} + \frac{s^2}{p_r} = \sum_{i=1}^{r} \frac{t_i^2}{p_i}, \tag{9}$$

donde

$$t_r = -s, \quad \sum_{i=1}^{r} t_i = 0. \tag{10}$$

Suponiendo $t = \hat{\theta}^* - p$ y notando que la condición (10) está cumplida, en calidad de (8) obtenemos

$$\sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i} > h_\varepsilon. \tag{11}$$

Esto no es otra cosa sino el criterio χ^2. De las afirmaciones citadas se deduce que $\chi^2(X) \in H_{r-1}$.

El criterio π'' en el teorema 13.2 equivale asintóticamente a (7) y (11) y tendrá la forma

$$\sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{\nu_i} > h_\varepsilon. \tag{12}$$

Teniendo también en cuenta el teorema 15.3 y la observación 15.1, podemos resumir lo dicho en la forma de la afirmación siguiente.

Teorema 1. *El criterio* (7) *para* $c_1 = h_\varepsilon/2$, *así como el criterio* χ^2 (11) *y el criterio* (12) *tienen un nivel asintótico* $1 - \varepsilon$ *y son los c.a.b. para verificar, basándose en la muestra* $X \in \mathrm{B}_\theta$, *la hipótesis* $\{\theta = p\}$ *frente a* $\{\theta \neq p\}$. *Estos son, a su vez, los criterios asintóticamente minimax para verificar la hipótesis* $\{\theta = p\}$ *frente a la alternativa* $\left\{\sum_{i=1}^{r}(\theta_i - p_i)^2/p_i > b^2/n\right\}$ *para cualquier* $b > 0$.

La equivalencia asintótica de los criterios (7), (11) y (12) también podría ser establecida directamente, utilizando el desarrollo en serie de $\ln \frac{\nu_i}{np_i} = \ln\left(1 + \frac{\nu_i - np_i}{np_i}\right)$ en (7).

Estos criterios son asintóticamente no paramétricos, ya que la distribución límite de las estadísticas que se utilizan en ellos es "absoluta", o sea, no está de ningún modo relacionada con la naturaleza de la distribución inicial.

2. Aplicaciones del criterio χ^2. Verificación de las hipótesis por los datos agrupados. El criterio χ^2 está ampliamente difundido y su importancia sale fuera de los límites del problema examinado en el apartado anterior.

Volvamos a examinar el problema general concerniente a la hipótesis $H_1 = \{X \in \mathbf{P}_1\}$ frente a $H_2 = \{X \in \mathbf{P}, \mathbf{P} \neq \mathbf{P}_1\}$ que hemos estudiado en el § 12. Puesto que la teoría de los criterios óptimos se ha desarrollado, en cierta medida, sólo en el caso paramétrico, es natural que se trate de "parametrizar" de algún modo este problema*).

En el caso general, la manera más simple y natural de hacer esto es la *agrupación de los datos*, que consiste en lo siguiente. El campo de los valores posibles de las magnitudes sujetas a observación (o sea, el espacio $\mathscr{X}$) se divide en r regiones disjuntas $\Delta_1, \ldots, \Delta_r$, y en vez de la observación x_j sólo se indica el intervalo Δ_k donde esta observación ha ido a parar.

*) Se tiene en cuenta un parámetro de dimensión finita. Cualquier problema puede considerarse paramétrica si se admite un parámetro de dimensión infinita, ya que éste puede ser identificado con la distribución $\mathbf{P}$, $X \in \mathbf{P}$.

Con otras palabras, reducimos la precisión de las observaciones, y los x_j que cayeron en Δ_k pueden ser sustituidos por un solo valor $z_k \in \Delta_k$. Claro está que eligiendo una división bastante completa, podemos aproximar la observación x_j mediante z_k tan exactamente como se quiera.

Así pues, la agrupación conduce a que la observación x_j es sustituida por el vector e_k si se ha producido el suceso $A_k = \{x_j \in \Delta_k\}$ (los vectores e_k han sido definidos al principio del apartado anterior). Pero la nueva muestra obtenida como resultado de tal operación, evidentemente, no es otra cosa sino la muestra de $\mathbf{B}_\theta$, $\theta_k = \mathbf{P}(x_j \in \Delta_k)$. Ya sabemos que en este caso, el vector $\nu = (\nu_1, \ldots, \nu_r)$ de frecuencias de caídas en los intervalos $\Delta_1, \ldots, \Delta_r$ será una estadística suficiente.

La reducción realizada de la muestra X al vector ν es precisamente la llamada agrupación de los datos.

Por supuesto que tal agrupación está relacionada con cierto "empobrecimiento" de la muestra X y con una pérdida parcial de información.

La parametrización realizada también puede ser considerada desde otro punto de vista. Supongamos, para evidenciar, que $\mathscr{X} = R$ y que todas las distribuciones que han de ser estudiadas, están concentradas en un intervalo finito y tienen densidad, o sea, satisfacen la condición (A_μ), donde μ es la medida de Lebesgue. Con la partición $\Delta_1, \ldots, \Delta_r$ establecida, examinemos, a la par con la densidad $f(x)$, la densidad constante a trozos

$$f_\theta(x) = \frac{\mathbf{P}(\Delta_i)}{\Delta_i} = \frac{1}{\Delta_i} \int_{\Delta_i} f(x)dx = \frac{\theta_i}{\Delta_i} \text{ para } x \in \Delta_i. \tag{13}$$

Donde Δ_i también designa la longitud del intervalo Δ_i. Esta es la familia paramétrica de las distribuciones $\mathbf{P}_\theta$, $\mathbf{P}_\theta(B) = \int_B f_\theta(x)dx$.

La muestra Y de $\mathbf{P}_\theta$ podrá ser obtenida si para cada k recogemos todas las observaciones de $X \in \mathbf{P}$ que han ido a parar a Δ_k y luego las "dispersamos" por Δ_k uniformemente y al azar. En realidad esto es lo mismo que hemos hecho antes, ya que los datos que indican en qué punto del intervalo Δ_k se encuentra la observación y_i, no contienen ninguna información acerca del parámetro θ: la función de verosimilitud $f_\theta(Y)$ no cambia después del "desplazamiento" de las observaciones dentro de los límites de sus intervalos. Por lo tanto, sólo es suficiente saber las cantidades $\nu_1, \ldots, \nu_r$ de observaciones que fueron a parar a $\Delta_1, \ldots, \Delta_r$.

Está claro que si $f(x)$ es una función suave, $f_\theta(x)$ aproximará bien $f(x)$ siempre que la partición de $\{\Delta_1, \ldots, \Delta_r\}$ sea bastante "menuda".

Las relaciones (13) significan otro método de parametrización, equivalente al primero. Tal equivalencia resulta de la coincidencia de las funciones de verosimilitud, con una exactitud de hasta un factor que no depende del

parámetro. Para la distribución (13), dicha equivalencia es igual a

$$f_\theta(Y) = \prod_{i=1}^{r} \theta_i^{\nu_i} \prod_{i=1}^{r} \Delta_i^{-\nu_i},$$

donde el primer factor es la función de verosimilitud de la muestra de $\mathbf{B}_\theta$ (véase (4)).

Cabe señalar que la agrupación de las observaciones a menudo también surge por sí misma no para fines de parametrización, sino simplemente como un método cómodo y económico de anotación de la información que contiene la muestra. Si, por ejemplo, $n = 10^4$ y la precisión de las mediciones de los valores observados en $[0, 1]$ es comparable con 0,1, entonces claro está que prácticamente no merece la pena conocer todas la 10^4 observaciones y es suficiente indicar 10 frecuencias $\nu_1, \ldots, \nu_{10}$ de caída en los intervalos $\Delta_i = ((i-1)/10, i/10)$, $i = 1, \ldots, 10$, o sea, basta conocer tan sólo el histograma de la muestra.

Volvamos al problema de verificación de la hipótesis $H_1 = \{X \in \mathbf{P}_1\}$ frente a $H_2 = \{X \in \mathbf{P} \neq \mathbf{P}_1\}$. Supondremos que la referida agrupación de observaciones es tal que la desviación (importante para nosotros) de la distribución $\mathbf{P}$ de la muestra X respecto a $\mathbf{P}_1$ se reflejará obligatoriamente en las distribuciones de los datos agrupados. Entonces, nuestro problema se puede considerar como un problema de verificación de la hipótesis $\{\theta = p\}$, donde $p_i = \mathbf{P}_1(\Delta_i)$, frente a $\{\theta \neq p\}$, para las familias paramétricas $\mathbf{B}_\theta$ o (13). Como ya sabemos, *en este problema, el criterio χ^2 (al igual que los criterios* (7) *y* (12)) *será asintóticamente óptimo desde el punto de vista enunciado en el teorema* 1.

Además, el criterio χ^2 *no es asintóticamente paramétrico*, ya que, para la hipótesis H_1, la distribución límite de la estadística $\chi^2(X)$ no depende de la distribución inicial de la muestra X.

En este caso cabe señalar que la verificación de la hipótesis $\{\theta = p\}$ para las familias (13) o $\mathbf{B}_\theta$ no es, a pesar de todo, equivalente a la verificación de la hipótesis $\{X \in \mathbf{P}_1\}$, aunque, con una partición abundante de $\{\Delta_1, \ldots, \Delta_r\}$, ella pueda ser próxima a esta última. En efecto, para la muestra X se verifica la hipótesis $X \in \mathbf{P}$, $\mathbf{P}(\Delta_i) = p_i = \mathbf{P}_1(\Delta_i)$. Esto contribuye a que el criterio χ^2 sea inconciliable respecto a las alternativas $\mathbf{P} \neq \mathbf{P}_1$ para las cuales $\theta_i = \mathbf{P}(\Delta_i) = \mathbf{P}_1(\Delta_i) = p_i$. Por eso indicaremos una vez más, que el criterio χ^2 es un criterio que posee una serie de propiedades de optimización asintótica, pero que actúa exclusivamente contra las alternativas que modifican el vector θ, o sea, contra las alternativas para las cuales $\{\mathbf{P}(\Delta_i)\} \neq \{\mathbf{P}_1(\Delta_i)\} = \{p_i\}$.

Hagamos algunas observaciones concernientes a las aplicaciones de los criterios χ^2, (7) y (12). En este caso hablaremos fundamentalmente tan sólo del criterio χ^2, ya que, por un lado, dichos criterios se asemejan unos a

otros y, por otro lado, el criterio χ^2 históricamente (en parte, debido a su evidencia) adquirió una aplicación mucho más amplia.

El nivel de importancia del criterio $\chi^2(X) > h_\varepsilon$ es igual a $1 - \varepsilon$ únicamente en el "límite". La experiencia muestra que para $\varepsilon \geqslant 0{,}01$, el verdadero nivel de importancia de este criterio se aproxima satisfactoriamente, mediante el valor de $1 - \varepsilon$, sólo cuando $np_i \geqslant 8$, $i = 1, \ldots, r$.

Si el número de grupos r es grande, digamos, cuando $n > r > 30$, se puede utilizar la aproximación normal tanto para la distribución $\frac{1}{\sqrt{2r}}(\chi_r^2 - r)$, $\chi_r^2 \in \mathbf{H}_r$ (véase el § 2.2), como también, en caso de la hipótesis H_1, para la distribución de la estadística $\chi^2(X)$ normalizada por los momentos

$$\mathbf{M}\chi^2(X) = r - 1,$$

$$\mathbf{D}\chi^2(X) = 2(r-1) + \frac{1}{n}\left(\sum_{i=1}^{r} p_i^{-1} - r^2 - 2r + 2\right).$$

Con frecuencia también se utiliza la aproximación normal $\Phi_{0,1}$ para distribuir la variable aleatoria (véase el § 2.2) $\sqrt{2\chi_r^2} - \sqrt{2r-1}$, $\chi_r^2 \in \mathbf{H}_r$.

También debemos señalar que al aumentar el número de grupos mejora la aproximación de la densidad $f(x)$ mediante una función escalonada construida según los valores de $\mathbf{P}_1(\Delta_i) = \int_{\Delta_i} f(x)dx$. Esto significa que aumenta el número de alternativas que no concuerdan con H_1, y que el criterio χ^2 se transforma cada vez más en criterio de verificación de las *hipótesis acerca de la densidad*. De acuerdo con esto, al aumentar el número de grupos, la potencia de los criterios χ^2 de nivel registrado disminuirá (compárese con las observaciones del párrafo anterior acerca del criterio de Morán. Esto se analiza más detalladamente en [12] y [21]).

Como defecto del criterio χ^2 debe considerarse el hecho de que en una serie de casos de partición $\{\Delta_1, \ldots, \Delta_r\}$ hay que establecer la estadística. Aquí es necesario tener cuidado, ya que en este caso se introduce un elemento de subjetivismo en el "empobrecimiento" de la muestra X. Además, a veces esta partición se elige en función de la muestra X, lo cual, hablando en general, no siempre es admisible, ya que, a su vez, Δ_i se vuelven aleatorias (esto se examina más detalladamente en [49], p. 575).

Ejemplo 1[*] En la ciudad N, un individuo observó las indicaciones de 500 relojes expuestos en las vitrinas de distintas relojerías. Los resultados de las observaciones fueron divididos en 12 grupos (conforme a la posición del horario en la esfera). He aquí la tabla de las observaciones obtenidas:

[*] Este ejemplo se ha tomado de [25].

Intervalos en la esfera	0—1	1—2	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12
Número de observaciones	41	34	54	39	49	45	41	33	37	41	47	39

Se verifica una hipótesis simple: H_1 = {*la distribución de la posición del horario en la esfera según los grupos de horas es uniforme*} frente a la alternativa adicional compuesta.

En este ejemplo, $n = 500$, $p_i = 1/12$, $i = 1, \ldots, 12$, $np_i \approx 41{,}67$. A base del teorema 1 podemos considerar que $\chi^2(X) \in \mathbf{H}_{11}$ aproximadamente. Sin embargo, en nuestro ejemplo, mediante el cálculo directo nos convencemos de que $\chi^2(X) \approx 10$, y el nivel realmente alcanzado del criterio χ^2 es aproximadamente igual a $1 - H_{11}((10, \infty)) \approx 0{,}47$ (véase la tabla III). Esto significa que los resultados del experimento concuerdan con la hipótesis H_1 desde el punto de vista del criterio χ^2 de cualquier nivel $1 - \varepsilon$ situado entre 0,47 y 1.

Ya hemos señalado que el criterio χ^2 está muy difundido. Además, la esfera de su aplicación consiste no sólo en verificar las hipótesis simples. Uno de tales ejemplos será examinado en el párrafo siguiente.

§ 17. Verificación de las hipótesis de pertenencia de la muestra a una familia paramétrica

Examinemos el problema de verificación de la hipótesis compuesta $H_1 = \{X \in \mathbf{P}_\alpha, \alpha \in A\}$ de que la distribución de la muestra pertenece a la familia paramétrica $\{\mathbf{P}_\alpha\}_{\alpha \in A}$ frente a la alternativa adicional $H_2 = \{X \in \mathbf{P}, \mathbf{P} \notin \{\mathbf{P}_\alpha\}_{\alpha \in A}\}$. Como ejemplo de tal género de hipótesis puede servir la afirmación de que X es la muestra de cualquier población normal (hipótesis H_1), así como la afirmación adicional a la mencionada (hipótesis H_2).

Como un segundo ejemplo puede servir la verificación de la hipótesis de que $X \in \mathbf{B}_{\theta(\alpha)}$, donde la dimensión de α es menor que la de θ. Este problema también puede ser interpretado como el problema de verificación de la hipótesis de pertenencia de X a una subfamilia paramétrica (véase el § 15). No obstante, la primera interpretación también será cierta, puesto que en el caso en que como resultado del experimento sólo acontezca un número finito de sucesos posibles (véase la definición de $\mathbf{B}_\theta$ en el § 2.2), la familia $\mathbf{B}_\theta$ comprenderá todas las distribuciones posibles de la muestra.

En el apartado siguiente examinaremos el problema de verificación de la hipótesis $X \in \mathbf{B}_{\theta(\alpha)}$ y mostraremos que el problema general de pertenencia a la familia paramétrica puede ser reducido al primer problema mediante la agrupación de los datos.

1. Verificación de la hipótesis $X \in \mathbf{B}_{\theta(\alpha)}$. Agrupación de los datos. Examinemos primeramente el problema general enunciado al principio del párrafo y destinado al espacio arbitrario $\mathscr{X}$. Dividamos el espacio $\mathscr{X}$ en regiones ("intervalos") $\{\Delta_1, ..., \Delta_r\}$ de tal modo que el número de "intervalos" r sea mayor que $l + 1$, donde l es la dimensión del parámetro α. Realicemos la agrupación de las observaciones en estos intervalos. Si la hipótesis $H_1 = \{X \in \mathbf{P}_\alpha\}$ es cierta, las probabilidades de que las observaciones caigan en los intervalos Δ_i serán iguales a

$$p_i(\alpha) = \mathbf{P}_\alpha(\Delta_i).$$

Esto significa que en este caso el vector $\theta = (\theta_1, ..., \theta_r)$ de las probabilidades de que las observaciones caigan en Δ_i debe situarse en la curva $\theta = p(\alpha) = (p_1(\alpha), ..., p_r(\alpha))$.

Ahora bien, a base de la muestra $Y \in \mathbf{B}_\theta$ obtenida en la agrupación, debemos verificar la hipótesis H_1 acerca de la pertenencia de Y a la subfamilia paramétrica $\mathbf{B}_{\theta(\alpha)}$ frente a la alternativa $\{Y \in \mathbf{B}_\theta\}$, donde θ no se sitúa en la curva $\theta = p(\alpha)$, $\alpha \in A$. Este problema fue examinado en el § 15, donde hemos hallado el criterio asintóticamente minimax para verificar H_1 frente a la alternativa semejante

$$H_2 = \{Y \in \mathbf{B}_\theta, \inf_{\theta} |\theta - p(\alpha_0 + \gamma n^{-1/2})I^{1/2}(p(\alpha_0 + \gamma n^{-1/2}))| > bn^{-1/2}\} \quad (1)$$

(veáse la aclaración 15.3 al teorema 15.4. El punto α_0 significa el valor "localizado" del parámetro, tal que las alternativas se disponen en el entorno del punto $\theta_0 = p(\alpha_0)$). En nuestro caso, el criterio de la relación de verosimilitud (15.11) tiene la forma

$$\ln R_1(X) = \max_\theta \sum_{i=1}^{r} \nu_i \ln \theta_i - \max_\alpha \sum \nu_i \ln p_i(\alpha) > h_\varepsilon/2,$$

o bien, que es lo mismo,

$$\sum_{i=1}^{r} \nu_i \ln \frac{\nu_i}{np_i(\hat{\alpha}^*)} > h_\varepsilon/2,$$

donde $\hat{\alpha}^*$ es la e.v.m. del parámetro α según la muestra Y o según el vector $\nu = (\nu_1, ..., \nu_r)$). Este criterio equivale asintóticamente (véase el teorema 15.4) al criterio

$$(p(\hat{\alpha}^*) - \nu n^{-1})I(p(\hat{\alpha}^*))(p(\hat{\alpha}^*) - \nu_n^{-1})^r > h_\varepsilon.$$

Como la forma de la matriz $I(\theta)$ es conocida (véase (16.5), entonces, utilizando (16.9), del teorema 15.4 obtenemos el

Corolario 1. *Si $r - 1 > l$ y la función $p(\alpha)$ satisface las condiciones del teorema 15.4, entonces el criterio de la relación de verosimilitud de nivel asintótico $1 - \varepsilon$ para verificar, basándose en los datos agrupados, la hipótesis $H_1 = \{X \in \mathbf{P}_\alpha, \mathbf{P}_\alpha \in \{\mathbf{P}_\alpha\}_{\alpha \in A}\}$ frente a la alternativa adicional H_2, es*

asintóticamente minimax (para verificar la hipótesis H_1 frente a (1)) y tiene la forma

$$\sum_{i=1} \nu_i \ln \frac{\nu_i}{np_i(\hat{\alpha}^*)} > h_\varepsilon/2, \tag{2}$$

donde h_ε es una cuantila del orden de $1 - \varepsilon$ de la distribución χ^2 con $r - l - 1$ grados de libertad. Este criterio equivale asintóticamente al criterio

$$\hat{\chi}^2(X) = \sum_{i=1}^{r} \frac{(\nu_i - np_i)(\hat{\alpha}_i^*))^2}{np_i(\hat{\alpha}_i^*)} > h_\varepsilon. \tag{3}$$

Este último criterio también se llama *criterio* χ^2, pero en caso de que los parámetros "obstaculizadores" desconocidos se estimen con arreglo a la muestra. Como se deduce del corolario 1, la distribución de la estadística $\hat{\chi}^2(X)$ converge, siempre que se trate de la hipótesis H_1, a la distribución χ^2 con $r - l - 1$ grados de libertad (el número de grados de libertad $r - 1$ en la distribución límite de la estadística $\chi^2(X)$ ha disminuido en el número de parámetros escalares $\alpha_1, \ldots, \alpha_l$ que se estiman por la muestra).

Ejemplo 1. En el ejemplo 2.26.3 hemos descrito el mecanismo de herencia de los grupos de sangre 0 (cero), A, B y AB. Este mecanismo es controlado por los genes de tres tipos A, B y 0. Las probabilidades de aparición de estos genes en una populación dada designémoslas por p, q, $r = 1 - p - q$. En el ejemplo 2.26.3 hemos hallado y en la tabla 1 del § 26 hemos escrito las probabilidades $p_i(\alpha)$ de que una persona tenga el i-ésimo grupo de sangre.

Disponemos de la muestra X con las frecuencias ν_i, $i = 1, 2, 3, 4$ (véase la tabla 1) de aparición del i-ésimo grupo de sangre, obtenida como resultado del examen de $n = 353$ personas. En el ejemplo 2.26.3 hemos hallado, para esta muestra, los valores aproximados de la e.v.m. $\hat{\alpha}^* = (p^*, q^*) = (0{,}246,\ 0{,}173)$. Esto nos proporcionó los valores de $p_i(\hat{\alpha}^*)$ expuestos en la tabla 1.

Tabla 1. Distribución de las personas según los grupos de sangre

	0	A	B	AB	En total
ν_i	121	120	79	33	353
$p_i^* = \nu_i n^{-1}$	0,343	0,340	0,224	0,093	1
$p_i(\hat{\alpha}^*)$	0,337	0,347	0,231	0,085	1

Hemos recibido la posibilidad de utilizar el corolario 1 para verificar la hipótesis acerca de que tiene lugar el mecanismo de herencia de la sangre, descrito anteriormente. Con ayuda de la tabla 1 hallamos que, en nuestro caso, la estadística $\hat{\chi}^2(X)$ (véase (3)) es igual a 0,44, aproximadamente. Esto concuerda bien con la hipótesis, ya que el valor crítico h_ε, correspondiente a la distribución χ^2 con un grado de libertad y al valor de $\varepsilon = 0{,}2$, es igual a $h_{0,2} \approx 1{,}64$.

Ejemplo 2. *Problema acerca de los indicios conjugados.* Supongamos que la muestra X es el resultado de la investigación de ciertos objetos, cada uno de los cuales se caracteriza por dos indicios A y B. El primero puede adoptar los valores $A_1, \ldots, A_s$, y el segundo, $B_1, \ldots, B_t$. Se pregunta, ¿serán esos indicios dependientes o no? Por ejemplo, podemos realizar cierto experimento G, obteniendo resultados $B_1, \ldots, B_t$ en condiciones $A_1, \ldots, A_s$ diferentes. El problema consiste en aclarar si dichos resultados dependen o no de las condiciones en que se realiza el experimento.

Este problema también puede considerarse como el problema de verificación de la independencia de dos variables aleatorias ξ y η según las observaciones agrupadas en el par (ξ, η).

En nuestro ejemplo, los resultados de los experimentos son una matriz de valores $\|\nu_{ij}\|$, donde ν_{ij} es el número de aparición de resultados con indicios A_i y B_j en la muestra X de volumen n (cada elemento de la muestra es un par de indicios del objeto que se examina).

Designemos $p_{ij} = \mathbf{P}(A_iB_j)$, $p_{i\cdot} = \sum_{j=1}^{t} p_{ij}$, $p_{\cdot j} = \sum_{i=1}^{s} p_{ij}$.

Entonces, la hipótesis H_1 de independencia de los indicios tendrá la forma $H_1 = \{p_{ij} = p_{i\cdot}p_{\cdot j}\}$. No es difícil notar que ésta es la hipótesis de pertenencia de la distribución de la muestra a una familia paramétrica, donde el papel de parámetro α lo desempeña el vector $\alpha = (p_{1\cdot}, \ldots, p_{s-1,\cdot}, p_{\cdot 1}, \ldots$ $\ldots, p_{\cdot t-1})$ de $s + t - 2$ dimensiones $\left(\text{los valores de } p_{s\cdot} \text{ y } p_{\cdot t} \text{ se deducen de las igualdades } p_{s\cdot} = 1 - \sum_{i=1}^{s-1} p_{i\cdot},\ p_{\cdot t} = 1 - \sum_{j=1}^{t-1} p_{\cdot j}\right)$.

La función de verosimilitud de la muestra X, siempre que se trate de la hipótesis H_1, es igual a

$$\prod_{i,j} p_{ij}^{\nu_{ij}} = \prod_i p_{i\cdot}^{\nu_{i\cdot}} \prod_j p_{\cdot j}^{\nu_{\cdot j}}, \quad \nu_{i\cdot} = \sum_{j=1}^{t} \nu_{ij}, \quad \nu_{\cdot j} = \sum_{i=1}^{s} \nu_{ij}.$$

De los resultados del § 16 (compárense con los del apartado (16.1)) se deduce que la e.v.m. $\hat{\alpha}^*$ para tal función de verosimilitud tiene la forma

$$p_{i\cdot}^* = \nu_{i\cdot}/n, \quad \hat{p}_{\cdot j}^* = \nu_{\cdot j}/n.$$

Así pues, en nuestro caso, el criterio χ^2 adquiere la forma

$$\hat{\chi}^2(X) = \sum_{i,j} \frac{(\nu_{ij} - n\hat{p}_{i.}^{*}\hat{p}_{.j}^{*})^2}{n\hat{p}_{i.}^{*}\hat{p}_{.j}^{*}} = n \sum_{i,j} \frac{(\nu_{ij} - n^{-1}\nu_{i.}\nu_{.j})^2}{\nu_{i.}\nu_{.j}} > h_\varepsilon,$$

donde h_ε es una cuantila del orden de $1 - \varepsilon$ de la distribución χ^2 con un número de grados de libertad de $st - 1 - (s + t - 2) = (s - 1)(t - 1)$.

Se pueden señalar muchos problemas aplicados, donde se utiliza el criterio de conjugación de los indicios que hemos construido. A título de ejemplo examinarenos uno de ellos: el problema de investigación sociológica de la relación entre los ingresos de las familias y la cantidad de niños en ellas (véase [25], p. 481).

Ejemplo 2A. Supongamos que el indicio A significa la cantidad de niños y adopta los valores 0, 1, 2, 3, $\geqslant 4$. El indicio B indica a qué intervalo $(0 - 1)$, $(1 - 2)$, $(2 - 3)$, $(\geqslant 3)$ (por unidad se han adoptado 1000 coronas suecas) pertenece el salario. Según los resultados de $n = 25\,263$ investigaciones se han obtenido los datos expuestos en la tabla 2.

Tabla 2

A \ B	0—1	1—2	2—3	$\geqslant 3$	ν_i
0	2161	3577	2184	1636	9558
1	2755	5081	2222	1052	11110
2	936	1753	640	306	3635
3	225	419	96	38	778
$\geqslant 4$	39	98	31	14	182
$\nu_{.j}$	6116	10928	5173	3016	25263

En este ejemplo, $\hat{\chi}^2(X) = 568{,}5$, lo cual supera en mucho el valor crítico de h_ε para la distribución χ^2 de $(5 - 1)(4 - 1) = 12$ grados de libertad, incluso con valores de ε bastante pequeños. Así que debemos reconocer la inconciliabilidad de la hipótesis $H_1 = \{A$ y B *son independientes* (*inconjugados*)$\}$.

No obstante, debemos señalar que un análisis más minucioso ha demostrado la existencia de una dependencia muy débil entre los indicios A y B.

2. Caso general. El criterio χ^2 aplicado al problema de este párrafo posee los mismos defectos que los indicados con arreglo a los problemas del párrafo anterior.

El problema de verificación de la hipótesis $\{X \in \mathbf{P}_\theta\}$ acerca de la pertenencia de X a la familia paramétrica $\{\mathbf{P}_\theta\}_{\theta \in \Theta}$ también admite, por supues-

to, un enfoque más amplio, análogo al expuesto en el § 12. Elijamos cierta distancia $d(\mathbf{P}, \mathbf{Q})$ en el espacio de distribuciones. Luego, hallemos el punto $\mathbf{P}_{\theta^*}$ de $\{\mathbf{P}_\theta\}$, inmediato a $\mathbf{P}_n^*$ desde el punto de vista de la distancia d. En calidad de $\mathbf{P}_{\theta^*}$ también se puede tomar $\mathbf{P}_{\hat\theta^*}$, donde $\hat\theta^*$ es la e.v.m. (véase el § 2.5) o cualquier otra estimación razonable. Si la hipótesis H_1 es cierta, entonces $d(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*)$ no debe ser grande y, al contrario, si es cierta H_2, entonces $d(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*)$ será considerable. Esta consideración nos ofrece la siguiente estructura del criterio: rechazaremos la hipótesis H_1 si $d(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*) > c$, y la aceptaremos en el caso contrario.

El número c debe elegirse de modo que

$$\sup_{\theta\in\Theta} \mathbf{P}_\theta(d(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*) > c) \leqslant \varepsilon,$$

o de modo que esta relación se cumpla asintóticamente. El corolario 1 propone que en calidad de distancia $d(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*)$ se adopten las estadísticas en (2) y (3). Entre otras, estas últimas también poseen la ventaja de que asintóticamente no son paramétricas: en el caso de la hipótesis $H_1 = \{X \in \mathbf{P}_\theta\}$, la distribución límite $\hat\chi^2(X)$ no depende, por ejemplo, de θ.

Examinemos la realización del enfoque general expuesto anteriormente en dos casos particulares importantes, cuando las familias paramétricas están formadas por parámetros de desplazamiento y de escala.

1) Supongamos que se verifica la hipótesis $X \in \mathbf{P}_\theta$, $\theta \in R$, donde $\mathbf{P}_\theta(A) = \mathbf{P}(A - \theta)$, $A \subset R$. Designemos por $F(x)$ la función de distribución correspondiente a $\mathbf{P}$ y pongamos $F_\theta(x) = F(x - \theta)$. En calidad de d adoptaremos la distancia que hemos utilizado en el criterio de Kolmogórov.

Teorema 1. *Supongamos que $X \in \mathbf{P}_\theta$, $F_\theta(x) = F(x - \theta)$ y que la función $F(x)$ tiene una densidad uniformemente continua limitada igual a $f(x) = F'(x)$, $\int x^2 f(x)dx < \infty$. Si designamos $\int xf(x)dx = a$, $\theta^* = \bar{x} - a$, entonces, cualquier θ*

$$\lim_{n\to\infty} \mathbf{P}_\theta(\sup_x \sqrt{n}|F_n^*(x) - F_{\theta^*}(x)| > c) = \mathbf{P}(\sup_x |w^\circ F(x)) + f(x)\int w^\circ(F(t))dt| > c,$$

donde w° es el puente browniano estándar.

En esta relación, el segundo miembro no depende de θ. Calculándolo para un valor dado de F y escogiendo $c = c_\varepsilon$ de modo que sea igual a ε, obtenemos el criterio

$$D_n = \sup_x \sqrt{n}|F_n^*(x) - F(x - \theta^*)| > c_\varepsilon$$

de nivel asintótico $1 - \varepsilon$ para verificar la hipótesis H_1 de pertenencia de la muestra X a la familia paramétrica $\{\mathbf{P}_\theta\}$, donde θ es el parámetro de desplazamiento.

Demostración del teorema 1. Examinemos el proceso

$$W_n(x) = \sqrt{n}(F_n^*(x) - F_{\theta^*}(x)) = w_n(x) - \sqrt{n}(F_{\theta^*}(x) - F_\theta(x)),$$

donde $w_n(x) = \sqrt{n}(F_n^*(x) - F_\theta(x))$. Para $t \to \theta$ tenemos

$$F_t(x) - F_\theta(x) = -(t - \theta)(f(x - \theta) + \varepsilon(t, \theta, x)),$$
$$|\varepsilon(t, \theta, x)| \leqslant \omega_{|t-\theta|},$$

donde ω_Δ es el módulo de continuidad de la función f, el cual no depende de x, $\omega_\Delta \to 0$ cuando $\Delta \to 0$. Como $\theta^* \underset{P_\theta}{\to} \theta$, entonces, poniendo $t = \theta^*$ y adoptando, sin limitar la generali-

dad, $a = 0$, obtenemos

$$\sqrt{n}(F_{\theta^*}(x) - F_\theta(x)) = -f(x-\theta)\int t\,d[\sqrt{n}(F_n^*(t) - F_\theta(t))] + \varepsilon(\theta^*, \theta, x) =$$
$$= -f(x-\theta)\int t\,dw_n(t) + \varepsilon(\theta^*, \theta, x),$$
$$|\varepsilon(\theta^*, \theta, x)| \leqslant \omega(\theta^* - \theta) = \sqrt{n}|\theta^* - \theta|\omega_{|\theta^* - \theta|} \underset{P_\theta}{\to} 0.$$

Seguidamente, la funcional

$$H_N(w_n) = \sup_x |w_n(x) - f(x-\theta)\int_{-N}^{N} w_n(t)dt|,$$

para cualquier $N > 0$, es continua en la métrica uniforme. La sustitución de la variable x por $F_\theta^{-1}(y) = \theta + F^{-1}(y)$, cuya ejecución natural para la aplicación del teorema 1.6.3, no modifica este hecho. Por eso, en virtud del teorema mencionado,

$$H_N(w_n) \Rightarrow \sup_x \left| w^\circ(F(x) - \theta)) + f(x-\theta)\int_{-N}^{N} w^\circ(F(t-\theta))dt \right|.$$

Para demostrar la relación requerida,

$$D_n \Rightarrow \sup_x |w^\circ(F(x-\theta)) + f(x-\theta)\int w^\circ(F(t-\theta))dt|$$

(el desplazamiento en θ del argumento en el segundo miembro no modifica el valor de este último) y, en virtud de las relaciones

$$|D_n - H_N(w_n) \leqslant \omega(\theta^* - \theta) + c\int_{|t| \geqslant N} w_n(t)dt \tag{4}$$
$$\omega(\theta^* - \theta) \underset{P_\theta}{\to} 0,$$

sólo queda convencernos que la integral en (4), juntamente con la integral $\int_{|t| \geqslant N} w^\circ(F(t)dt$ (pongamos, para abreviar, $\theta = 0$), convergen, de modo probable, a cero cuando $n \to \infty$, $N \to \infty$. Por lo visto, el método más simple de estimar ambas integrales consiste en demostrar la pequeñez de sus dispersiones utilizando la desigualdad de Chébishev. En vista de que los primeros dos momentos de las expresiones subintegrales en ambas integrales se comportan del mismo modo, podemos limitarnos a estimar tan sólo una de estas últimas. Examinemos, por ejemplo,

$$\int_{-\infty}^{-N} w^\circ(F(t))dt.$$

En virtud de las relaciones $\mathbf{M}w^\circ(s)w^\circ(u) = \min(s, u) + su \leqslant 2\min(s, u)$ cuando $s \leqslant 1$ y $u \leqslant 1$, tenemos

$$\mathbf{M}\left(\int_{\infty}^{-N} w^\circ(F(t))dt\right)^2 \leqslant 2\int_{-\infty}^{-N}\int_{-\infty}^{-N} \min(F(t), F(s))dtds =$$
$$= 4\int_{-\infty}^{-N}(-t-N)F(t)dt \leqslant -8\int_{-\infty}^{-N} tF(t)dt \to 0$$

cuando $N \to \infty$, ya que $\int t^2\,dF(t) < \infty$. Análogamente se examinan los demás intervalos. ◁

2) Supongamos que ahora se verifica la hipótesis $X \in \mathbf{P}_\theta$, $\theta \in R$, $\theta > 0$, donde $\mathbf{P}_\theta(A) = \mathbf{P}(A/\theta)$, $A \subset R$. Volvamos a designar por F la función de distribución correspondiente a $\mathbf{P}_1$ y pongamos

$$\sigma^2 = \mathbf{M}_1 x_1^2 = \int x^2 \mathbf{P}(dx), \quad \theta^* = \frac{1}{\sigma}\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}.$$

Teorema 2. *Supongamos que* $X \in \mathbf{P}_\theta$, $F_\theta(x) = F(x/\theta)$ *y que existe una densidad continua limitada* $f(x) = F'(x)$ *tal, que*

$$\sup_x |xf(x)| < \infty, \quad \int x^4 f(x)dx < \infty. \tag{5}$$

Entonces, para cualquier θ,

$$\lim_{n\to\infty} \mathbf{P}_\theta\left(\sup_x \sqrt{n}|F_n^*(x) - F(x/\theta^*)| > c\right) =$$
$$= \mathbf{P}\left(\sup_x |w^\circ(F(x)) + xf(x)\int tw^\circ(F(t))dt| c\right).$$

La demostración de este teorema es absolutamente análoga a la del teorema 1. Tenemos

$$W_n(x) = \sqrt{n}(F_n^*(x) - F(x/\theta^*)) = w_n(x) - \sqrt{n}(F(x/\theta^*) - F(x/\theta)),$$
$$w_n(x) = \sqrt{n}(F_n^*(x) - F(x/\theta)).$$

Cuando $t \to \theta$

$$F_t(x) - F_\theta(x) = \left(\frac{x}{t} - \frac{x}{\theta}\right)\cdot\left(f\left(\frac{x}{\theta}\right)\right) + \varepsilon(t, \theta, x),$$

donde, en virtud de la relación $f(x) < c/|x|$ y de la continuidad uniforme de f en cualquier intervalo finito, se cumple $\sup_x |\varepsilon(t, \theta, x)| \leqslant \omega_{|t-\theta|} \to 0$. Poniendo $t = \theta^* \underset{\mathbf{P}_\theta}{\to} \theta$, obtenemos

$$\sqrt{n}\left(F\left(\frac{x}{\theta^*}\right) - F\left(\frac{x}{\theta}\right)\right) =$$
$$= \sqrt{n}\left(\frac{x}{\theta^*} - \frac{x}{\theta}\right) f\left(\frac{x}{\theta}\right) - \frac{\sqrt{n}(\theta^* - \theta)}{\theta^*}\cdot\frac{x}{\theta}\cdot f\left(\frac{x}{\theta}\right)\varepsilon(\theta^*, \theta, x),$$

donde $\sup_x$ del segundo sumando converge a cero respecto a la $\mathbf{P}_\theta$-probabilidad. Sólo nos queda utilizar los razonamientos del teorema anterior (la pequeñez de las integrales $\int_{|t|>N} tw^\circ(F(t))dt$ y $\int_{|t|>N} tw_n(t)dt$ es asegurada por la condición (5)) y señalar que la parte principal $W_n(x)$ es igual a (adoptemos, sin limitar la generalidad, $\sigma^2 = 1$)

$$w_n(x) - \frac{\sqrt{n}\, x(\theta^{*2} - \theta^2)}{\theta\theta^*(\theta + \theta^*)} f(x/\theta) = w_n(x) - \frac{xf(x/\theta)}{\theta\theta^*(\theta + \theta^*)}\int t^2\, dw_n(t) =$$
$$= w_n(x) - \frac{2xf(x/\theta)}{\theta\theta^*(\theta + \theta^*)}\int tw_n(t)dt,$$

donde $\theta^*(\theta + \theta^*) \underset{P_\theta}{\to} 2\theta^2$. Por eso

$$\sup_x |W_n(x)| = \sup_x \left| w^\circ\left(F\left(\frac{x}{\theta}\right)\right) + \theta^{-3} x f\left(\frac{x}{\theta}\right) \int t w^\circ\left(F\left(\frac{t}{\theta}\right)\right) dt \right| =$$

$$= \sup_x \left| w^\circ\left(F\left(\frac{x}{\theta}\right)\right) + \frac{x}{\theta} f\left(\frac{x}{\theta}\right) \int t w^\circ(F(t)) dt \right|.$$

Como la transformación de la contracción respecto a x debajo del signo $\sup_x$ no modifica nada, el teorema 2 queda demostrado.

El lector también puede obtener resultados análogos para las estadísticas $\int (F_n^*(x) - F_{\theta^*}(x))^2 dF_\theta(x)$.

§ 18. Estabilidad de las decisiones estadísticas

Al construir distintos procedimientos estadísticos en los párrafos anteriores — en los problemas de estimación o de verificación de las hipótesis — cada vez partíamos de cierto conjunto de condiciones. Estas últimas se referían, en particular, a la independencia de las observaciones y a su igual distribución, así como a las suposiciones acerca del carácter de distribución **P** de los elementos de una muestra. El incumplimiento de tales condiciones significaría que las afirmaciones respectivas (por ejemplo, acerca del carácter de distribución límite o acerca de la optimización de una u otra estadística) son, hablando en general, inciertas.

Por otro lado, en la práctica, las referidas condiciones son, como regla, el resultado de la aproximación y la idealización inevitable. Por consiguiente, dichas condiciones suelen no cumplirse de manera exacta y surgen dudas acerca de la validez de las recomendaciones basadas en uno u otro procedimiento estadístico elegido.

Por lo tanto, al igual que en cualquier otra rama de las matemáticas, referente a las aplicaciones, aquí es necesario (en la última etapa, antes de aplicar los métodos elaborados) aclarar cuán grandes deben ser las divergencias de las condiciones adoptadas, para que este hecho nos obligue a modificar las conclusiones enunciadas.

Desde el punto de vista matemático, tal procedimiento constituye un problema muy parecido al problema de la estabilidad. En los libros editados en inglés, para este tipo de problemas se ha adoptado el término "robustness"*). Por eso en los manuales editados en ruso, a la par con el término "estabilidad" también se utiliza la palabra "robusticidad".

Las divergencias más difundidas de las condiciones antes mencionadas consisten en lo siguiente.

*) Robustez o robusticidad.

1) En la serie de observaciones X está presente una pequeña porción de "desechos", o sea, de observaciones provocadas por graves errores de medición o de registro, o engendradas por cualquier otro mecanismo "obstaculizador", distinto del sistema sujeto a investigación. Por lo general, la separación de dichas observaciones es imposible. En vez de esto se buscan procedimientos que sean poco sensibles a tal "ensuciamiento" de la muestra.

2) La distribución de x_i no equivale con exactitud a $\mathbf{P}$, sino que tan sólo aproximadamente.

3) Los elementos de la muestra X no son independientes, sino que tan sólo débilmente dependientes.

La tarea consiste en construir las reglas resolutivas para los problemas principales de la estadística matemática, que sean semejantes, por su eficacia, a las reglas óptimas y que al mismo tiempo sean insensibles a las referidas divergencias de las condiciones adoptadas o, al menos, a aquellas de ellas que para nosotros no tienen importancia. Esta tarea, dificilísima y no siempre planteada con exactitud, aún no está estudiada del todo. Aquí los resultados tienen un carácter muy heterogéneo. Por eso sólo nos detendremos en algunos ejemplos típicos.

1. Estimación de la media para las distribuciones simétricas. Supongamos que $X \in \mathbf{P}$ y que la distribución en la recta $\mathbf{P}$ tiene una densidad de $f(t-\alpha)$ respecto a la medida de Lebesgue, $f(t) = f(-t)$. Examinemos las dos estimaciones siguientes del parámetro $\alpha = \mathbf{M}x_1$. Una de ellas es

$$\alpha^* = \bar{x},$$

y la otra, α^{**}, que se basa en las cuantilas muestrales:

$$\alpha^{**} = \frac{1}{r-1} \sum_{k=1}^{r-1} \zeta^*_{kp}, \tag{1}$$

donde $0 < p < 1$, $r = 1/p$ es un número entero. Cuando $p = 1/2$, la estimación α^{**} se transforma en la mediana muestral $\zeta^* = \zeta^*_{1/2}$.

Limitémonos por ahora al caso de $p = 1/2$. Cuando $n \to \infty$ tenemos

$$(\alpha^* - \alpha)\sqrt{n} \in \Phi_{0,\sigma_1^2}, \quad \sigma_1^2 = \int t^2 f(t)\,dt. \tag{2}$$

Además, en el corolario 2.2.1 hemos establecido que para $n \to \infty$

$$(\alpha^{**} - \alpha)\sqrt{n} \in \Phi_{0,\sigma_2^2}, \quad \sigma_2^2 = \frac{1}{4f^2(0)}. \tag{3}$$

Analizando la demostración de este corolario es fácil establecer que junto con $\alpha^{**} = \zeta^* = x_{(k_0)}$, $k_0 = [(n+1)/2]$, esa misma distribución límite se observará en el término de la serie variacional $x_{(k)}$ para cualquier valor registrado de la diferencia $k - k_0$.

De aquí se deduce que la estimación $\alpha^{**} = \zeta^*$ es insensible (desde el punto de vista de sus propiedades asintóticas) al hecho de que a la muestra X se agregue cualquier número finito de "desechos". En efecto, si tenemos l "desechos" cualesquiera en la muestra X, entonces α^{**} se situará entre los valores $y_{(k_1)}$ e $y_{(k_2)}$, donde $k_1 = k_0 - l$, $k_2 = k_0 + l$ e $y_{(k)}$, $k = 1, \ldots$ $\ldots, n - l$ forman la serie variacional de la muestra $Y \in \mathbf{P}$ de volumen $n - l$. Pero las propiedades asintóticas de $y_{(k_1)}$ e $y_{(k_2)}$ son iguales y coinciden con las de la mediana muestral.

Así pues, cualesquiera que sean los "desechos", la estimación α^{**} será insensible a ellos. Eso no se puede decir de la estimación $\alpha^* = \bar{x}$, donde los referidos desechos pueden influir considerablemente (por ejemplo, si son comparables, en cuanto a su magnitud, con n). Es fácil comprender que la propiedad de estabilidad de α^{**} también se conservará para pequeñas muestras, si el número de desechos l es pequeño respecto a n. Asimismo esta propiedad se conservará en el caso en que en vez de ζ^* se utilice una estadística (1) de una forma más general.

Por otro lado, para un caso particular importante, cuando $\mathbf{P} = \Phi_{\alpha,\sigma_1^2}$, hay una ley normal: el valor de $\sigma_2^2 = \sigma_1^2\pi/2(f(0) = (\sigma_1\sqrt{2\pi})^{-1})$ excede la dispersión σ_1^2 de la estimación eficiente $\alpha^* = \bar{x}$ solamente $\pi/2$ veces. Esta diferencia entre la eficacia de α^{**} y α^* puede disminuir aún más si las estimaciones (1) se examinan cuando $r = 3, 4$, etc. Entonces obtendremos una estimación α^{**} casi tan eficiente como $\bar{x}$ (al carecer de desechos) y al mismo tiempo estable respecto a los desechos. Además de (1) se puede tomar la media truncada

$$\alpha^{**} = \frac{1}{n - 2np} \sum_{k = np+1}^{n - np} x_{(k)}, \tag{4}$$

cuya dispersión también se aproxima con pequeños valores de p) a la dispersión σ_1^2 de la estimación α^*.

Señalemos a continuación, que las propiedades de la estimación $\alpha^* = \bar{x}$ dependen poco de las variaciones de $\mathbf{P}$, que conservan la varianza $\sigma_1^2 = \int t^2 f(t)dt$ y, en particular, de las variaciones locales de $f(t)$ en el punto $t = 0$. En este sentido dicha estimación es estable. Pero su *propiedad de optimización*, que tiene lugar para $\mathbf{P} = \Phi_{\alpha,\sigma_1^2}$, es inestable. En efecto, supongamos que para un valor pequeño de $\varepsilon > 0$,

$$\mathbf{P} = (1 - \varepsilon)\Phi_{\alpha,1} + \varepsilon \mathbf{U}_{\alpha-\varepsilon,\alpha+\varepsilon}.$$

Entonces $f(0) = (1 - \varepsilon)/\sqrt{2\pi} + 1/2 > 1/2$ y, como muestran las relaciones (2) y (3), la estimación $\alpha^{**} = \zeta^*$ será mucho mejor (el valor de ε debe ser pequeño, pero no menor de $(1/\sqrt{n})$).

Por otro lado, la estimación $\alpha^{**} = \zeta^*$ es estable (se tiene en cuenta

su distribución) respecto a las variaciones de **P** que no afecten el valor de $f(0)$.

Las observaciones expuestas también pueden enunciarse de otro modo: con arreglo a los criterios estadísticos, por ejemplo, a los c.u.m.p. no desplazados $|\overline{x} - \alpha_0| > c$ para verificar, a partir de la muestra $X \in \Phi_{\alpha,1}$, la hipótesis $H_1 = \{\alpha = \alpha_0\}$ frente a $H_2 = \{|\alpha - \alpha_0| > d > 0\}$.

2. Estadística de Student y S_0^2. Examinemos ahora la cuestión concerniente a la estabilidad de los procedimientos estadísticos (estimación y verificación de las hipótesis) relacionados con las estadísticas

$$t = \frac{(\overline{x} - \alpha)\sqrt{n}}{S_0}, \qquad S_0^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

Como sabemos (véase los §§ 3.7 y 3.8), en estas estadísticas se basan los criterios óptimos para verificar, correspondientemente, las hipótesis respecto a la media α y a la varianza σ^2 de las poblaciones normales en el caso cuando se desconoce el segundo parámetro (σ^2 o α) de la distribución Φ_{α,σ^2}.

Las estadísticas t y S_0^2 se comportan de manera diferente con arreglo a las alteraciones de las condiciones $X \in \Phi_{\alpha,\sigma^2}$. Supongamos que n es grande y $X \in \mathbf{P}$, donde **P** es cualquier distribución, con α media y con varianza finita. Entonces, la distribución t, al igual que en el caso $X \in \Phi_{\alpha,\sigma^2}$, se aproximará a la distribución normal $\Phi_{0,1}$. Esto se deduce de los teoremas de continuidad (§ 1.5) y del hecho de que

$$(\overline{x} - \alpha)\sqrt{n}/\sqrt{\mathbf{D}x_1} \in \Phi_{0,1}, \quad S_0^2 \underset{p}{\to} \mathbf{D}_1.$$

Lo dicho significa que la dimensión del criterio de Student se diferenciará poco, para grandes valores de n, de la dimensión dada, si incluso la distribución **P** de la muestra X se diferencia considerablemente de la distribución normal.

Esto no se puede decir con arreglo a los criterios construidos a base de la estadística S_0^2. Esta circunstancia se debe al hecho de que la distribución límite S_0^2 depende del valor $\mathbf{M}x_1^4$. En efecto, de las consideraciones del capítulo 1 resulta

$$(S_0^2 - \sigma^2)\sqrt{n} \in \Phi_{0,d^2}, \quad d^2 = \mathbf{M}(x_1^2 - \sigma^2)^2 = \mathbf{D}x_1^2.$$

Por consiguiente, la dimensión del criterio construido a base de la estadística S_0^2 para una población normal puede diferenciarse considerablemente de la dimensión dada, si $X \in \mathbf{P}$ y **P** se diferencian de Φ_{α,σ^2} (pero si coinciden los cuartos momentos de **P** y Φ_{α,σ^2}, entonces no habrá diferencia).

Ambas estadísticas t y S_0^2 son sensibles al rechazamiento de la suposición acerca de la independencia de las observaciones en la muestra X. Si,

por ejemplo, todas las observaciones en la muestra están relacionadas unas con otras, y el coeficiente de correlación es igual a ϱ, entonces, adoptando $\alpha = 0$ sin limitar la generalidad, obtenemos

$$\mathbf{M}S_0^2 = \frac{1}{n-1}\mathbf{M}\left[\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2\right] =$$

$$= \frac{1}{n-1}\left[n\sigma^2 - \frac{1}{n}\mathbf{M}\left(\sum_{i=1}^{n} x_i\right)^2\right] =$$

$$= \frac{1}{n-1}\left[n\sigma^2 - \sigma^2(1-\varrho) - n\sigma^2\varrho\right] = \sigma^2(1-\varrho).$$

Ahora bien, aquí se altera incluso la propiedad de no desplazamiento de S_0^2, aunque para pequeños valores de ϱ la divergencia será pequeña. El establecimiento de las distribuciones de t y S_0^2 suele chocar con grandes dificultades al aparecer cierta dependencia.

3. **Criterio de relación de verosimilitud.** Este criterio suele ser muy sensible a la existencia de desechos e incluso de pequeñas divergencias en las suposiciones acerca de la distribución de X. Supongamos, por ejemplo, que se verifican dos hipótesis simples $H_1 = \{X \in \Phi_{0,1}\}$ y $H_2 = \{X \in \mathbf{U}_{-1,1}\}$. Está claro que, al utilizar el criterio más potente de Neyman — Pearson, la aparición incluso de una sola observación x fuera del segmento $[-1, 1]$, siempre que las demás observaciones correspondan idealmente a la distribución $\mathbf{U}_{-1,1}$, nos obligará (¡con una probabilidad nula de equivocación!) a reconocer la hipótesis H_1. Esto significa que la presencia de un solo desecho o la aparición incluso de pequeñas divergencias de la distribución $\mathbf{U}_{-1,1}$ pueden obligarnos a tomar una decisión falsa.

En este sentido, el criterio de Kolmogórov es, por ejemplo, mucho más estable (aunque también menos potente respecto a H_2). En general, los criterios no paramétricos, como era de esperar, son mucho más estables que los criterios "individuales" dotados de propiedades de optimización en uno u otro problema concreto.

En cuanto al referido problema de verificación de la normalidad (H_1) frente a la uniformidad (H_2) de la muestra X, el establecimiento de criterios potentes y al mismo tiempo estables respecto a los desechos, se puede realizar utilizando, como antes, la relación de verosimilitud, pero para muestras "truncadas" (compárese con (4)). También se puede ir por la vía de elección de otro criterio cualquiera. En este sentido, la existencia de una reserva bastante grande de criterios y estimaciones diferentes es muy útil. A esto a menudo se acude no sólo por razones de estabilidad, sino también por cuestiones de comodidad de los cálculos.

CAPÍTULO 4

Problemas estadísticos de dos muestras y más

En los §§ 1 y 2 se examinan los problemas de homogeneidad de dos muestras.

En el § 3 se estudian los problemas de regresión.

En el § 4 se exponen los resultados del análisis de varianza.

En el § 5 se examinan los problemas de reconocimiento de las imágenes.

§ 1. Verificación de las hipótesis de homogeneidad (completa o parcial) en el caso paramétrico

1. Clase de problemas a examinar. En los capítulos anteriores, el objeto de todos los estudios ha sido la muestra X de volumen n de una distribución $\mathbf{P}$ total o parcialmente desconocida. Ahora pasamos al estudio de los problemas estadísticos donde figura no una, *sino dos muestras y más.*

Una de las clases principales de problemas que se examinan en este caso son los problemas de verificación de la homogeneidad (completa o parcial) de dos muestras.

Aquí entran los tres siguientes tipos principales de problemas:

I. *Verificación de la homogeneidad "ordinaria".* Aquí el problema consiste en verificar la hipótesis de que dos muestras X e Y se han extraído de una misma distribución desconocida. Tales problemas surgen, por ejemplo, al comparar dos métodos de elaboración en cualquier proceso tecnológico o en la agricultura. Como base de comparación suelen servir las características numéricas del producto final (de la muestra), que son de naturaleza aleatoria. Problemas de este mismo género surgirán si por el estado de salud de los enfermos verificamos el efecto de una nueva medicina, comparando el grupo experimental de pacientes con el grupo de control.

Entre los problemas de homogeneidad figura el ejemplo dado en la introducción.

En este párrafo examinaremos el *caso paramétrico*. Supongamos que se da una familia de distribuciones $\{\mathbf{P}_\theta\}_{\theta \in \Theta}$ y que hay dos muestras independientes $X = (x_1, \ldots, x_{n_1})$ e $Y = (y_1, \ldots, y_{n_2})$ de volúmenes n_1 y n_2, respectivamente, con la particularidad de que se sabe de antemano que estas muestras pertenecen a la familia $\{\mathbf{P}_\theta\}$:

$$X \in \mathbf{P}_{\theta_1}, \qquad Y \in \mathbf{P}_{\theta_2} \tag{1}$$

para ciertos θ_1 y θ_2. El problema ordinario de homogeneidad aquí consiste en verificar la hipótesis $H_1 = \{\theta_1 = \theta_2\}$ frente a la alternativa adicional $H_2 = \{\theta_1 \neq \theta_2\}$. Es evidente que aquí ambas hipótesis H_1 y H_2 son compuestas.

II. *Verificación de la homogeneidad al existir un parámetro obstaculizador.* Aquí se supone que la dimensión k del parámetro θ es mayor que 1. Escribamos el vector θ en forma de la colección $\theta = (u, v)$ de dos subvectores u y v y designemos por u_j las componentes de los vectores θ_j en (1), $j = 1, 2$.

Supongamos que sabemos de antemano que en ambas muestras, "el subparámetro", a pesar de ser desconocido, es común: $v_1 = v_2 = v$. Se verifica la hipótesis $H_1 = \{u_1 = u_2\}$ frente a $H_2 = \{u_1 \neq u_2\}$.

Este es precisamente un problema de homogeneidad cuando se dispone del parámetro obstaculizador v. El mismo se distingue de los problemas ordinarios de homogeneidad por el hecho de que la alternativa para la hipótesis $H_1 = \{\theta_1 = \theta_2\}$ tiene la forma $H_2 = \{u_1 \neq u_2,\ v_1 = v_2\}$.

Se puede citar el siguiente ejemplo de surgimiento de tal tipo de problemas. Supongamos que nos interesa el estado de cierto objeto que se caracteriza por el vector a que no puede ser medido directamente. Podemos efectuar tan sólo mediciones en las que sobre a se superpone un ruido aleatorio cuya naturaleza, al efectuar diversas observaciones, permanece invariable. Debemos verificar la hipótesis de invariabilidad de a en dos series de observaciones X e Y.

Si, digamos, las mediciones tienen la forma $x_i = a_1 + \xi_i$, donde $\xi_i \in \Phi_{\lambda,\sigma^2}$ determinan el papel que desempeña el ruido, y las observaciones y_i tienen ese mismo carácter al sustituir a_1 por a_2, entonces podemos escribir $X \in \Phi_{a_1+\lambda,\sigma^2}$, $Y \in \Phi_{a_2+\lambda,\sigma^2}$. Hemos llegado al problema de verificación de la igualdad de las medias $\{\alpha_1 = \alpha_2\}$ de dos poblaciones normales Φ_{α_1,σ^2} y Φ_{α_2,σ^2} para el valor desconocido común σ^2.

III. *Verificación de la homogeneidad parcial.* Aquí solamente se verifica la hipótesis H_1 acerca de la coincidencia "parcial" de θ_1 y θ_2. Es decir, se comprueba la hipótesis $H_1 = \{u_1 = u_2\}$ (con designaciones del apartado anterior) frente a $H_2 = \{u_1 \neq u_2\}$. Los valores de v_1 y v_2 pueden ser propios para cada una de las muestras X e Y.

Supongamos, por ejemplo, que en un laboratorio se estima el resultado

de la influencia que ejerce un nuevo método de cultivo sobre el rendimiento de cualquier cereal. Las observaciones representan el peso total de los granos en distintas espigas. Supongamos que $x_i \in \Phi_{\alpha_1,\sigma_1^2}$, $i = 1, \dots, n_1$ para una partida experimental de espigas, e $y_i \in \Phi_{\alpha_2,\sigma_2^2}$ para la partida de control. Es natural admitir que la "dispersión" σ^2 puede variar a consecuencia del cambio de cultivo. Pero para nosotros es importante saber si cambia o no el índice principal α que determina el rendimiento del cereal. Llegamos al problema de verificación de la hipótesis $H_1 = \{\alpha_1 = \alpha_2\}$ frente a $H_2 = \{\alpha_1 \neq \alpha_2\}$ para poblaciones normales cuyas varianzas pueden ser diferentes. En la literatura, este problema es conocido con el nombre de *problema de Behrens — Fisher**).

En este párrafo reduciremos los problemas de todos los tres tipos, para las familias paramétricas arbitrarias, al problema examinado en el § 3.15, de pertenencia de una muestra a una subfamilia paramétrica, y hallaremos una serie de criterios asintóticamente minimax, suponiendo la semejanza de las hipótesis sometidas a verificación. Serán los criterios de relación de verosimilitud que, para poblaciones normales, coincidirán con los criterios construidos al buscar una u otra optimización exacta (si tales existen; compárese con [57]).

El criterio estadístico π para verificar H_1 frente a H_2, en nuestro caso será la función $\pi = \pi(X, Y)$ de dos muestras X e Y que, al igual que en la exposición anterior, designará la probabilidad de aceptación de H_2 para una muestra unida dada (X, Y) (véase el capítulo 3). Las definiciones del nivel asintótico y de la optimización asintótica del criterio π aquí son las mismas que en el § 3.14.

Definición 1. Diremos que el criterio π *tiene un nivel asintótico* $1 - \varepsilon$ (pertenece a la clase K_ε), si

$$\limsup_{n\to\infty} \sup_{(\theta_1,\theta_2)\in\overline{\Theta}_1} \mathbf{M}_{\theta_1\theta_2}\pi(X, Y) \leqslant \varepsilon_1,$$

donde $\mathbf{M}_{\theta_1\theta_2}$ significa la esperanza matemática respecto a la distribución $\mathbf{P}_{\theta_1} \times \mathbf{P}_{\theta_2}$, y $\overline{\Theta}_1$ es el conjunto de valores (θ_1, θ_2) con los que se cumple

*) Se han escrito muchos libros dedicados a la búsqueda de sus soluciones óptimas. Al estudio del problema de Behrens — Fisher, que resultó muy difícil, contribuyeron considerablemente las investigaciones de Yu. V. Línnik y sus alumnos. Dichas investigaciones requieren la introducción de nuevos conceptos y el uso de un aparato matemático muy complejo. Esto hace imposible la enunciación y demostración (en el marco de este manual) de los resultados obtenidos. La situación acerca de los problemas de homogeneidad ordinaria y de homogeneidad para poblaciones normales al existir un parámetro obstaculizador, es algo mejor (en una serie de problemas se logra hallar los criterios invariantes no desplazados y uniformemente más potentes). No obstante, las construcciones indispensables para ello también resultan muy complicadas; este tema se examina más detalladamente en [57].

la hipótesis H_1 (por ejemplo, el conjunto de todos los puntos (θ_1, θ_2) situados en la "bisectriz" $\theta_1 = \theta_2$ en el problema de homogeneidad ordinaria).

Definición 2. El criterio $\pi_1 \in \bar{K}_\varepsilon$ se llama asintóticamente minimax en $\bar{K}_\varepsilon$ para verificar H_1 frente a H_2, si para cualquier criterio $\pi \in \bar{K}_\varepsilon$ se cumple

$$\lim_{n\to\infty}\inf\Big(\inf_{(\theta_1,\theta_2)\in\bar{\Theta}_2} \mathbf{M}_{\theta_1\theta_2}\pi_1(X, Y) - \inf_{(\theta_1,\theta_2)\in\bar{\Theta}_2} \mathbf{M}_{\theta_1\theta_2}\pi(X, Y)\Big) \geqslant \theta_1,$$

donde $\bar{\Theta}_2$ es el conjunto de valores (θ_1, θ_2) correspondientes a las alternativas de H_2.

2. Criterio asintóticamente minimax para verificar las hipótesis semejantes de homogeneidad ordinaria. Introduzcamos un nuevo parámetro $\bar{\theta} = (\theta_1, \theta_2)$ que caracterice la muestra "unida" (X, Y). La función de verosimilitud de la muestra es igual a $f_{\bar{\theta}}(X, Y) = f_{\theta_1}(X)f_{\theta_2}(Y)$.

Supongamos primeramente, para abreviar, que los volúmenes de las muestras coinciden: $n_1 = n_2 = n$. Entonces, la muestra (X, Y) puede representarse como muestra de volumen n formada por las observaciones $(x_1, y_1), \ldots, (x_n, y_n)$ de la distribución $\mathbf{P}_{\bar{\theta}} = \mathbf{P}_{\theta_1} \times \mathbf{P}_{\theta_2}$ que tiene la densidad $f_{\theta_1}(x)f_{\theta_2}(y)$. Llegamos al problema examinado en el § 3.15, de verificación, a base de la muestra (X, Y), de la hipótesis H_1 de que el parámetro $\bar{\theta}$ se sitúa en la "curva" $\theta_1 = \theta_2$. Teniendo en cuenta las designaciones de § 3.15, en nuestro caso, la hipótesis H_1 tiene la forma $\bar{\theta} = g(\alpha)$, donde $\alpha = \theta_1$, $g(\alpha) = (\alpha, \alpha)$. Es evidente que la matriz $G = \left\| \frac{\partial g_i}{\partial \alpha_j} \right\|$, $i = 1, \ldots, 2k$, $j = 1, \ldots, k$, tiene la forma $\begin{pmatrix} E \\ E \end{pmatrix}$, donde E es la matriz unidad de k-ésimo orden, así que el rango de G es igual a k.

Consideraremos localizado el parámetro $\bar{\theta}$, o sea, consideraremos que los valores de θ_1 y θ_2 son semejantes y, por consiguiente, que los posibles valores de $\bar{\theta}$ se sitúan en el entorno del punto $\bar{\theta}_0 = (\theta_0, \theta_0)$ para cierto θ_0 registrado. Si seguimos el § 3.15, nos será más cómodo introducir un nuevo parámetro $\tau = (\tau', \tau'') = (\gamma'/\sqrt{n}, \gamma''/\sqrt{n}) = \gamma/\sqrt{n}$, donde $\tau' = \theta_1 - \theta_0$, $\tau'' = \theta_2 - \theta_1$, así que la aplicación $\bar{\theta} = \bar{\theta}(\tau)$ es biunívoca: $\theta_1 = \tau' + \theta_0$, $\theta_2 = \tau'' + \tau' + \theta_0$. En los términos de los parámetros τ y γ, la hipótesis H_1 de homogeneidad tomará la forma $H_1 = \{\tau'' = 0\} = \{\gamma'' = 0\}$. En calidad de alternativa examinaremos la hipótesis "aislada"

$$H_2^b = \{\gamma'' I \gamma''^T \geqslant b^2\}, \; b > 0, \tag{2}$$

donde $I = I(\theta_0)$ es la matriz de Fisher para la familia $\{\mathbf{P}_\theta\}$ en el punto θ_0.

Teorema 1. *Supongamos que en el entorno del punto* θ_0, *la familia* $\{\mathbf{P}_\theta\}$ *satisface las condiciones* (RR) (véase el § 2.28). *Entonces, el criterio de rela-*

ción de verosimilitud

$$R_1(X, Y) = \frac{\sup\limits_{\theta_1,\theta_2} f_{\theta_1}(X)f_{\theta_2}(Y)}{\sup\limits_{\theta} f_\theta(X)f_\theta(Y)} > e^{h_\varepsilon/2} \quad (3)$$

es el criterio asintóticamente minimax de nivel $1 - \varepsilon$ *para verificar* $H_1 = \{\theta_1 = \theta_2\}$ frente a $H_2^b = \{(\theta_1 - \theta_2)I(\theta_1 - \theta_2)^T \geqslant b^2/n\}$ *para cualquier* $b > 0$, *donde* h_ε *es una cuantila de orden* $1 - \varepsilon$ *de la distribución* χ^2 *de* k *grados de libertad (para la hipótesis* H_1, *la estadística* $2 \ln R_1(X, Y)$ *tiene tal distribución límite).*

Supongamos que $\hat{\theta}_X^*, \hat{\theta}_Y^*, \hat{\theta}^*$ *es la e.v.m. del parámetro* $\theta = \theta_1 = \theta_2$, *respectivamente, según las muestras* X, Y, (X, Y). *Entonces, el criterio*

$$(\hat{\theta}_X^* - \hat{\theta}^*)I(\hat{\theta}^*)(\hat{\theta}_X^* - \hat{\theta}^*)^T + (\hat{\theta}_Y^* - \hat{\theta}^*)I(\hat{\theta}_Y^* - \hat{\theta}^*)^T > h_\varepsilon/n \quad (4)$$

será asintóticamente equivalente al criterio (3).

Demostración. La afirmación mencionada es el corolario directo del teorema 3.15.4. Sólo debemos aclarar qué representa la matriz de Fisher $\bar{I}(\bar{\theta}_0) = \bar{I}(\theta_0, \theta_0)$ para el parámetro "unido" $\bar{\theta} = (\theta_1, \theta_2)$, y la matriz $\mathbf{M}_2$ para la familia paramétrica $\{\mathbf{P}_{(\theta_0, \theta_0 + \beta)}\}$ en el punto $\beta = 0$. Tenemos

$$\ln f_{\theta_1}(x)f_{\theta_2}(y) = l(x, \theta_1) + l(y, \theta_2).$$

Designemos por t_i, $i = 1, \ldots, 2k$ las coordenadas del vector $\bar{\theta}$. En este caso, si por $\mathbf{M}_{\bar{\theta}}$ se designa la esperanza matemática en la distribución $\mathbf{P}_{\bar{\theta}}$, los elementos $\bar{I}_{ij}(\bar{\theta})$ de matriz $\bar{I}(\theta)$ serán iguales a

$$\bar{I}_{ij}(\bar{\theta}) = \mathbf{M}_{\bar{\theta}} \left(\frac{\partial l(x_1, \theta_1)}{\partial t_i} + \frac{\partial l(y_1, \theta_2)}{\partial t_i} \right) \left(\frac{\partial l(x_1, \theta_1)}{\partial t_j} + \frac{\partial l(y_1, \theta_2)}{\partial t_j} \right).$$

De aquí, en virtud de la independencia de x_1 e y_1, obtenemos

$$\bar{I}(\bar{\theta}) = \begin{pmatrix} I(\theta_1) & 0 \\ 0 & I(\theta_2) \end{pmatrix}.$$

Por eso, el criterio (4) no es otra cosa sino el criterio (3.15.12) en el teorema 3.15.4.

Los cálculos análogos muestran que $\mathbf{M}_2 = I(\theta_0)$, ya que para $\beta = \beta(\beta_1, \ldots, \beta_k) = 0$,

$$\frac{\partial l(x_1, \theta_0)}{\partial \beta_i} + \frac{\partial l(y_1, \theta_0 + \beta)}{\partial \beta_i} = \frac{\partial l(y_1, \theta_0)}{\partial t_i}, \quad i = 1, \ldots, k. \ \triangleleft$$

Observación 1. La afirmación del teorema 1 se ha obtenido suponiendo que $n_1 = n_2$. Sin embargo, esta limitación no tiene absolutamente importancia. Examinemos, por ejemplo, el caso cuando $n_1 \to \infty$, $n_2 \to \infty$, de modo que la relación n_1/n_2 sea igual a un número racional r_1/r_2 (r_1 y r_2 son

números enteros arbitrarios registrados, $n_i = nr_i$, $n \to \infty$). Volvamos a introducir el nuevo parámetro $\bar{\theta} = (\theta_1, \theta_2)$ y examinemos la muestra unida (X, Y) como una muestra de volumen n con las observaciones $(x_1, \ldots, x_{r_1}; y_1, \ldots, y_{r_2}), (x_{r_1+1}, \ldots, x_{2r_1}; y_{r_2+1}, \ldots, y_{2r_2}), \ldots$ de la distribución

$$\mathbf{P}_{\bar{\theta}} = \underbrace{\mathbf{P}_{\theta_1} \times \ldots \times \mathbf{P}_{\theta_1}}_{r_1 \text{ veces}} \times \underbrace{\mathbf{P}_{\theta_2} \times \ldots \times \mathbf{P}_{\theta_2}}_{r_2 \text{ veces}}$$

que depende del parámetro $\bar{\theta}$. La función de verosimilitud otra vez adquirirá la forma

$$f_{\bar{\theta}}(X, Y) = f_{\theta_1}(X) f_{\theta_2}(Y).$$

Si se introduce, como antes, el nuevo parámetro $\tau = (\tau', \tau'') = (\theta_1 - \theta_0, \theta_2 - \theta_1)$ y se pone $\tau = \gamma/\sqrt{n} = (\gamma'/\sqrt{n}, \gamma'/\sqrt{n})$, entonces, el problema sometido a examen consiste en verificar $H_1 = \{\gamma'' = 0\}$ frente a $H_2^b = \{\gamma'' \mathbf{M}_2 \gamma''^T \geqslant b^2\}$, donde $\mathbf{M}_2$ es la matriz de Fisher para $\mathbf{P}_{(\theta_0, \theta_0 + \beta)}$ en el punto $\beta = 0$. Es fácil ver que en nuestro caso $\mathbf{M}_2 = r_2 I(\theta_0)$, así que el conjunto de alternativas conserva su forma (2):

$$H_2^b = \{\gamma'' I \gamma''^T > b^2/r^2\}.$$

La matriz de Fisher $I(\bar{\theta})$ tendrá la forma

$$\begin{pmatrix} r_1 I(\theta_1) & 0 \\ 0 & r_2 I(\theta_2) \end{pmatrix}.$$

Sólo queda utilizar el teorema 3.15.4. Entonces obtendremos la afirmación del teorema 1, en la que el criterio (4) ha de sustituirse por

$$n_1(\hat{\theta}_X^* - \theta^*) I(\theta^*)(\hat{\theta}_X^* - \theta^*)^T + \\ + n_2(\hat{\theta}_Y^* - \theta^*) I(\theta^*)(\hat{\theta}_Y^* - \theta^*) > h_\varepsilon. \quad (5)$$

Con ayuda del teorema 3.15.4 también se puede señalar la potencia asintótica garantizada de los criterios (3) — (5).

La afirmación del teorema también es válida en el caso general cuando $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to c$, donde c es un número arbitrario de $(0, 1)$. No obstante, la demostración de este hecho exige consideraciones adicionales.

Observación 2. La afirmación del teorema 1 también será válida si la hipótesis $H_1 = \{\theta_1 = \theta_2\}$ se sustituye por la hipótesis (véanse los capítulos precedentes)

$$H_1^a = \{(\theta_1 - \theta_2) I (\theta_1 - \theta_2)^T \leqslant a^2/n\}, \; 0 < a < b.$$

Observación 3. La forma de criterios asintóticamente minimax en el teorema 1 no depende de θ_0. El valor de θ_0 sólo forma parte de la definición de la hipótesis H_2^b a través de $I = I(\theta_0)$ (véase (2), aunque también sería posible evitar la aparición de θ_0 sustituyendo I en (2) por $I((\theta_1 + \theta_2)/2)$. Esto nos proporcionaría la hipótesis H_2^b ("asintóticamente equivalente" a

H^b_2), para la cual se conserva por completo la afirmación del teorema 3. La aparición del valor θ_0 en (2) se debe a la utilización del método más simple de reducción del referido problema a los resultados del § 3.15.

Ejemplo 1. Supongamos que X e Y son muestras de volúmenes n_1 y n_2 de las distribuciones polinomiales $X \in \mathbf{B}_{\theta_1}$, $Y \in \mathbf{B}_{\theta_2}$, $\theta_l \in R^k$, $\theta_l = (\theta_{l1}, \ldots, \theta_{lk})$, $l = 1, 2$. Los vectores de las frecuencias $\nu = (\nu_1, \ldots, \nu_k)$ y $\mu = (\mu_1, \ldots, \mu_k)$ de aparición de los sucesos $A_1, \ldots, A_k$ (véase el § 2.2) forman las estadísticas suficientes

$$f_{\theta_1}(X) = \prod_{i=1}^{k} \theta_{1i}^{\nu_i}, \quad f_{\theta_2}(Y) = \prod_{i=1}^{k} \theta_{2i}^{\mu_i}.$$

Las e.v.m. tienen la forma $\theta^*_X = \nu/n_1$, $\theta^*_Y = \mu/n_2$, $\theta^* = (\nu + \mu)/(n_1 + n_2)$. La matriz $I(\theta)$ está definida en (3.15.5), así que (veáse (3.16.9))

$$tI(\theta_0)t^T = \sum_{i=1}^{k} \frac{t_i^2}{\theta_{0i}}.$$

Así pues, en virtud del teorema 1 y de la observación 1, el criterio asintóticamente minimax de nivel asintótico $1 - \varepsilon$ para verificar $H_1 = \{\theta_1 = \theta_2\}$ frente a

$$H_2^b = \left\{ \sum_{i=1}^{k} (\theta_{1i} - \theta_{2i})^2/\theta_{0i} \geqslant b^2/n_2 \right\}$$

tiene la forma

$$\ln R_1(X, Y) =$$

$$= \sum_{i=1}^{k} \nu_i \ln \frac{\nu_i}{n_1} + \sum_{i=1}^{k} \mu_i \ln \frac{\mu_i}{n_2} - \sum (\nu_i + \mu_i) \ln \frac{\nu_i + \mu_i}{n_1 + n_2} > \frac{h_\varepsilon}{2},$$

donde h_ε es una cuantila del orden de $1 - \varepsilon$ de la distribución χ^2 con $k - 1$ grados de libertad. De acuerdo con (4) y (5), será asintóticamente equivalente el criterio

$$n_1 \sum_{i=1}^{k} \left(\frac{\nu_i}{n_1} - \frac{\nu_i + \mu_i}{n_1 + n_2} \right)^2 \frac{n_1 + n_2}{\nu_i + \mu_i} +$$

$$+ n_2 \sum_{i=1}^{k} \left(\frac{\mu_i}{n_2} - \frac{\nu_i + \mu_i}{n_1 + n_2} \right)^2 \frac{n_1 + n_2}{\nu_i + \mu_i} =$$

$$= \sum_{i=1}^{k} \left(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2} \right)^2 \frac{n_1 n_2}{\nu_i + \mu_i} > h_\varepsilon. \quad (6)$$

Ejemplo 1A. En el ejemplo 2.26.3 hemos descrito el mecanismo de herencia de los grupos de sangre designados por 0 (cero), A, B y AB. Dicho mecanismo es controlado por genes de tres tipos: A, B y 0. Las probabilidades de que esos genes aparezcan en una población dada se designan por p, q, $r = 1 - p - q$, respectivamente. Las probabilidades $p_i(\alpha)$, $\alpha = (p, q)$ de que una persona tenga el i-ésimo grupo de sangre se expresan a través de α según las fórmulas citadas en la tabla 1 del § 2.26.

Tenemos dos muestras X e Y con frecuencias ν_i y μ_i, $i = 1, \ldots, 4$ de aparición del i-ésimo grupo sanguíneo, obtenidas a consecuencia del examen de $n_1 = 353$ personas de la comunidad I, de $n_2 = 364$ personas de la comunidad II. La distribución de las personas según los grupos sanguíneos se da en la tabla 1

Tabla 1

	0	A	B	AB	Total
Comunidad I	121	120	79	33	353
Comunidad II	118	95	121	30	364
Total	239	215	200	63	717

Es necesario verificar la hipótesis de pertenencia de las comunidades examinadas a una población, o sea, la hipótesis de igualdad de las probabilidades p y q de estos grupos o, que es lo mismo, la hipótesis de igualdad de las probabilidades $p_i(\alpha)$. Éste es, evidentemente, el problema de homogeneidad examinado en el ejemplo 1.

Si se verifica la coincidencia de las probabilidades de los cuatro grupos de sangre, entonces, a la estadística (veánse los capítulos precedentes)

$$\chi_1^2 = \sum_{i=1}^{4} \left(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2} \right)^2 \frac{n_1 n_2}{\nu_i + \mu_i}$$

le corresponderá la distribución χ^2 con tres grados de libertad. En nuestro caso el valor χ_1^2 constituye 11,74. El nivel realmente alcanzable (véase el § 3.4) de la desviación obtenida pasa de 0,99. Esto significa que la hipótesis de homogeneidad ha de ser rechazada desde el punto de vista del criterio $\chi_1^2 > h_{0,01}$ de nivel 0,99.

Debemos señalar que el criterio aplicado no del todo corresponde a la naturaleza del fenómeno examinado, ya que debemos verificar la coincidencia de las probabilidades p y q y no la de las probabilidades p_i de aparición de los grupos sanguíneos. Ateniéndose exactamente al teorema 1,

debemos, mediante los métodos descritos en el § 2.26, calcular las e.v.m. $\hat{\alpha}_X^*$, $\hat{\alpha}_Y^*$ y $\hat{\alpha}^*$ del parámetro $\alpha = (p, q)$ con arreglo a las muestras X, Y y (X, Y), respectivamente, y utilizar la estadística

$$\chi_2^2 = 2[L(\hat{\alpha}_X^*, X) + L(\hat{\alpha}_Y^*, Y) - L(\hat{\alpha}^*, (X, Y))] =$$
$$= 2\left[\sum_{l=1}^{4} \nu_l \ln p_l(\hat{\alpha}_X^*) + \sum_{l=1}^{4} \mu_l \ln p_l(\hat{\alpha}_Y^*) - \sum_{l=1}^{4} (\nu_l + \mu_l) \ln p_l(\hat{\alpha}^*)\right]$$

que tiene, con grandes valores de n, una distribución próxima a la distribución χ^2 con dos grados de libertad. Si realizamos todos los cálculos necesarios (véase el ejemplo 2.26.3), obtendremos $\chi_2^2 \approx 11{,}04$, lo cual proporciona, para dos grados de libertad, una desviación mayor de 11,74 para tres grados de libertad.

En cuanto a la verificación de la propia hipótesis de pertenencia de X e Y a las subfamilias paramétricas $\mathbf{B}_{p(\alpha)}$, donde $p(\alpha) = p_1(\alpha), \ldots, p_4(\alpha)$, véase el ejemplo 3.17.1. Ambas muestras concuerdan bien con esta hipótesis.

Ejemplo 2. Sea $X \in \Phi_{\alpha_1,\sigma_1^2}$, $Y \in \Phi_{\alpha_2,\sigma_2^2}$, donde los puntos $\theta_l = (\alpha_l, \sigma_l^2)$ se sitúan en el entorno del punto $\theta_0 = (\alpha_0, \sigma_0^2)$. Aquí

$$I(\theta_0) = \begin{pmatrix} \sigma_0^{-2} & 0 \\ 0 & \frac{1}{2}\sigma_0^{-4} \end{pmatrix}$$

(véase el § 2.16), y examinaremos el problema de verificación de la hipótesis $H_1 = \{\theta_1 = \theta_2\}$ frente a

$$H_2^\delta = \left\{ \frac{(\alpha_1 - \alpha_2)^2}{\sigma_0^2} \frac{(\sigma_1^2 - \sigma_2^2)^2}{2\sigma_0^4} \geqslant \frac{b^2}{n_2} \right\}, \quad n = n_1 + n_2.$$

Tenemos $\hat{\theta}_X^* = (\bar{x}, S_X^2)$, $S_X^2 = \dfrac{1}{n_1}\displaystyle\sum_{l=1}^{n_1} (x_l - \bar{x})^2$, $f_{\hat{\theta}_X^*}(X) = (2\pi e S_X^2)^{-n_1/2}$.

Las fórmulas análogas son válidas para la muestra Y. Seguidamente

$$\hat{\theta}^* = (\bar{z}, S_{X,Y}^2), \ \bar{z} = \frac{\left(\sum_{l=1}^{n_1} x_l + \sum_{l=1}^{n_2} y_l\right)}{n_1 + n_2} = a\bar{x} + (1 - a)\bar{y}, \tag{7}$$

$$S_{X,Y}^2 = \frac{1}{n_1 + n_2}\left[\sum_{l=1}^{n_1} (x_l - \bar{z})^2 + \sum_{l=1}^{n_2} (y_l - \bar{z})^2\right] =$$
$$= aS_X^2 = (1 - a)S_Y^2 + (1 - a)a(\bar{x} - \bar{y})^2,$$

donde $a = n_1/(n_1 + n_2)$, $f_{\hat{\theta}}(X)f_{\hat{\theta}}(Y) = (2\pi e S^2_{X,Y})^{-\frac{1}{2}(n_1+n_2)}$. Ahora bien, para verificar H_1 frente a H_2^ε, como criterio asintóticamente minimax utilizaremos el criterio

$$\frac{S^2_{X,Y}}{S_X^{2a} S_Y^{2(1-a)}} > e^{h_\varepsilon/(n_1+n_2)},$$

donde h_ε es la cuantila de la distribución χ^2 con dos grados de libertad. Le proponemos al lector que halle, en calidad de ejercicio, el criterio asintóticamente equivalente que tiene la forma (5).

3. Criterios asintóticamente minimax para el problema de homogeneidad al existir un parámetro obstaculizador. En éste y en los apartados posteriores supondremos, para abreviar, que los volúmenes de las muestras X e Y coinciden: $n_1 = n_2$. Esta limitación no tiene importancia. En el caso de $n_1/n_2 = r_1/r_2$ (r_1 y r_2 son enteros) el lector puede liberarse por sí mismo de esta limitación así como se hizo en la observación 1 del teorema 1.

Así pues, supongamos que se dan dos muestras $X \in \mathbf{P}_{\theta_1}$ e $Y \in \mathbf{P}_{\theta_2}$, $\theta_i = (u_i, v_i)$, $i = 1, 2$, de volúmenes $n_1 = n_2 = n$. Se verifica la hipótesis $\{u_1 = u_2\}$ frente a $\{u_1 \neq u_2\}$ suponiendo que conocemos $v_1 = v_2 = v$ y v. La dimensión u_i se designa por l, $l < k$.

Introduzcamos un nuevo parámetro $\bar{\theta} = (u_1, u_2, v)$. Representemos la muestra unida (X, Y) como una muestra de volumen n con observaciones $(x_1, y_1), \ldots, (x_n, y_n)$ cuya densidad de distribución es igual a $f_{\bar{\theta}}(x, y) = f_{(u_1, v)}(x)f_{(u_2, v)}(y)$. Para esta familia paramétrica, el problema sometido a investigación equivale al problema de verificación de la hipótesis H_1, que consiste en el hecho de que el valor de $\bar{\theta}$ se encuentra en la "curva" $\bar{\theta} = g(\theta_1) = (u_1, u_1, v)$ frente a la alternativa adicional. La matriz $G = \left\| \frac{\partial g_i}{\partial \theta_{1j}} \right\|$, $i = 1, \ldots, k + 1$, $j = 1, \ldots, k$, tiene la forma $\begin{pmatrix} E_l & 0 \\ & E_k \end{pmatrix}$, donde arriba se halla la matriz unidad de orden l, y abajo, la matriz unidad de orden k, así que el rango de G es igual a k.

Al igual que en el apartado anterior, consideraremos que el parámetro θ ha sido localizado cerca del punto $\theta_0 = (u_0, v_0)$. Introduzcamos el parámetro $\tau = \tau(\bar{\theta}) = (\tau', \tau'', \tau''') = (u_1 - u_0, u_2 - u_1, v - v_0)$. La aplicación inversa $\bar{\theta} = \bar{\theta}(\tau)$ siempre existe y sus coordenadas son $u_1 = \tau' + u_0$, $u_2 = \tau'' + \tau' + u_0$, $v = \tau''' + v_0$. Pongamos $\tau = \gamma/\sqrt{n}$, $\gamma = (\gamma', \gamma'', \gamma''')$.

Para el nuevo parámetro τ(o γ), la hipótesis de homogeneidad tiene la forma $H_1 = \{\gamma'' = 0\}$. En calidad de alternativa examinemos la hipótesis "aislada" $H_2^\varepsilon = \{\gamma'' I_1(\theta_0)\gamma^T \geqslant b^2\}$, donde $I_1(\theta)$ es la submatriz de la matriz inicial de información de Fisher $I(\theta)$, formada por sus primeras l filas y columnas.

Teorema 2. *Supongamos que en el entorno del punto θ_0, la familia $\{\mathbf{P}_\theta\}$ satisface las condiciones (RR). Entonces, el criterio de relación de verosi-*

militud

$$R_1(X, Y) = \frac{\sup_{(u_1, u_2, v)} f_{(u_1, v)}(X) f_{(u_2, v)}(Y)}{\sup_\theta f_\theta(X) f_\theta(Y)} > e^{h_\varepsilon/2} \tag{8}$$

es el criterio asintóticamente minimax, de nivel asintótico $1 - \varepsilon$, *para verificar* $H_1 = \{u_1 = u_2\}$ *frente a*

$$H_2^b = \{(u_1 - u_2) I_1(\theta_0)(u_1 - u_2)^T \geqslant b^2/n\}, \tag{9}$$

con un valor común de $v_1 = v_2 = v$ *y con cualquier* $b > 0$. *Aquí* h_ε *es una cuantila del orden de* $1 - \varepsilon$ *de la distribución* χ^2 *con* l *grados de libertad. (Tal será la distribución límite* $2 \ln R_1(X, Y)$ *en la hipótesis* H_1).

Designemos por $\bar{\theta}^*$ *el valor del parámetro* $\bar{\theta}$ *con el que se alcanza el valor máximo del numerador* en (8), y por $\theta^* = (u^*, v^*)$, *el valor de* θ *con el que se alcanza el valor máximo del denominador. Representemos la matriz* $I(\theta)$ *en la forma*

$$I(\theta) = \begin{pmatrix} I_1(\theta) & I_{21}(\theta) \\ I_{12}(\theta) & I_{22}(\theta) \end{pmatrix}.$$

Entonces, el criterio

$$(\bar{\theta}^* - (u^*, u^*, v^*))\bar{I}(\bar{\theta}^*)(\bar{\theta}^* - (u^*, u^*, v^*))^T > h_\varepsilon/n, \tag{10}$$

donde

$$\bar{I}(\bar{\theta}) = \begin{pmatrix} I_1(\theta_1) & 0 & I_{21}(\theta_1) \\ 0 & I_1(\theta_2) & I_{21}(\theta_2) \\ I_{12}(\theta_1) & I_{12}(\theta_2) & I_{22}(\theta_1) + I_{22}(\theta_2) \end{pmatrix}, \tag{11}$$

será asintóticamente equivalente a (8).

Demostración. Este teorema también es el corolario directo del teorema 3.15.4. Sólo queda aclarar la estructura de la matriz $\bar{I}(\bar{\theta})$ para la muestra (X, Y) del parámetro "unido" $\bar{\theta}$ y de la matriz $\mathbf{M}_2$. Tenemos

$$l \equiv \ln f_{\bar{\theta}}(x, y) = l(x, (u_1, v)) + l(y, (u_2, v)).$$

Designemos por t_i, $i = 1, \ldots, k + l$, las coordenadas del vector $\bar{\theta}$. Entonces

$$\frac{\partial l}{\partial t_i} = \begin{cases} \dfrac{\partial l(x, (u_1, v))}{\partial t_i}, & 0 < i \leqslant l, \\ \dfrac{\partial l(y, (u_2, v))}{\partial t_i}, & l < i \leqslant 2l, \\ \dfrac{\partial l(x, (u_1, v))}{\partial t_i} + \dfrac{\partial l(y, (u_2, v))}{\partial t_i}, & 2l < i \leqslant k + l; \end{cases}$$

de aquí se obtiene (11) sin dificultad.

La matriz $\mathbf{M}_2$ para la familia paramétrica $\mathbf{P}_{\bar{\theta}(0,\beta,0)} = \mathbf{P}_{(u_0,\, u_0+\beta,\, v_0)}$ en el punto $\beta = 0$ se calcula análogamente. La misma es igual a $I_1(\theta_0)$ y corresponde a la submatriz media de la matriz $\bar{I}(\bar{\theta}_0)$. ◁

En los ejemplos expuestos consideraremos que los volúmenes de las muestras n_1 y n_2 son arbitrarios.

Ejemplo 3. Sea $X \in \Phi_{\alpha_1, \sigma^2}$, $Y \in \Phi_{\alpha_2, \sigma^2}$. Es necesario verificar la hipótesis $H_1 = \{\alpha_1 = \alpha_2\}$ cuando se desconoce σ^2. Para determinar los criterios asintóticamente minimax con ayuda del teorema 2 necesitamos hallar la estadística $R_1(X, Y)$ en (8), donde en nuestro caso $u_1 = \alpha_i$, $v = \sigma^2$, $\bar{\theta} = (\alpha_1, \alpha_2, \sigma^2)$. Tenemos $\ln f_{(\alpha_1, \sigma^2)}(X) f_{(\alpha_2, \sigma^2)}(Y) = -\frac{1}{2}(n_1 + n_2) \ln (2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n_1} (x_i - \alpha_1)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n_2} (y_i - \alpha_2)^2$. Reduciendo a cero las derivadas de esta función respecto a α_1, α_2 y σ^2, y resolviendo las ecuaciones obtenidas, hallamos según las designaciones del ejemplo 2)

$$\bar{\theta}^* = (\bar{x}, \bar{y}, aS_X^2 + (1 - a)S_Y^2), \quad a = \frac{n_1}{n_1 + n_2}, \tag{12}$$

$$f_{\bar{\theta}^*}(X, Y) = [2\pi e(aS_X^2 + (1 - a)S_Y^2)]^{-(n_1+n_2)/2}.$$

Procediendo del mismo modo con la función $\ln f_\theta(X) f_\theta(Y) = \ln f_{(\alpha,\sigma^2)}(X) f_{(\alpha,\sigma^2)}(Y)$, obtenemos (véase el ejemplo 2)

$$\theta^* = (\bar{z}, S_{X,Y}^2),$$

$$f_{\theta^*}(X) f_{\theta^*}(Y) = (2\pi e S_{X,Y}^2)^{-\frac{1}{2}(n_1+n_2)}. \tag{13}$$

Ahora bien, el criterio asintóticamente óptimo tiene la forma

$$\frac{S_{X,Y}^2}{aS_X^2 + (1 - a)S_Y^2} > e^{h_\varepsilon/(n_1+n_2)}$$

o bien (veáse (7))

$$\frac{\sqrt{a(1 - a)}|\bar{x} - \bar{y}|}{\sqrt{aS_X^2 + (1 - a)S_Y^2}} > \sqrt{\frac{h_\varepsilon}{n_1 + n_2}},$$

donde h_ε es una cuantila del orden de $1 - \varepsilon$ de la distribución χ^2 con un solo grado de libertad, así que $\sqrt{h_\varepsilon}$ se puede sustituir por el valor de $\lambda_{\varepsilon/2}$ para el cual $\Phi_{0,1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) = 1 - \varepsilon$. Es fácil notar que el primer miembro de la desigualdad

$$\frac{\sqrt{a(1 - a)(n_1 + n_2)}|\bar{x} - \bar{y}|}{\sqrt{aS_X^2 + (1 - a)S_Y^2}} > \lambda_{\varepsilon/2} \tag{14}$$

que define el criterio asintóticamente minimax, después de sustituir $|\bar{x} - \bar{y}|$ por $\bar{x} - \bar{y}$ será asintóticamente normal con los parámetros (0, 1) de una variable aleatoria.

Pero este criterio puede ser exacto (o sea, puede tener con exactitud un nivel dado de antemano). Efectivamente, en virtud de los resultados del § 2.32, en el caso de la hipótesis H_1,

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\bar{x} - \bar{y}}{\sigma} \in \Phi_{0,1},$$

$$\frac{(n_1 + n_2) a S_X^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 \in \mathbf{H}_{n_1 - 1},$$

$$\frac{(n_1 + n_2)(1 - a) S_Y^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n_2} (y_i - \bar{y})^2 \in \mathbf{H}_{n_2 - 1}.$$

En vista de que las tres variables aleatorias son independientes, la relación

$$(\bar{x} - \bar{y}) \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left[\frac{n_1 + n_2}{n_1 + n_2 - 2} (a S_X^2 + (1 - a) S_Y^2) \right]^{-1/2} =$$

$$= \frac{(\bar{x} - \bar{y}) \sqrt{a(1 - a)(n_1 + n_2 - 2)}}{\sqrt{a S_X^2 + (1 - a) S_Y^2}} \in \mathbf{T}_{n_1 + n_2 - 2}$$

tiene distribución de Student con $n_1 + n_2 - 2$ grados de libertad. Así pues, el criterio (compárese con (14))

$$\frac{(\bar{x} - \bar{y}) \sqrt{a(1 - a)(n_1 + n_2 - 2)}}{\sqrt{a S_X^2 + (1 - a) S_Y^2}} > \tau_\varepsilon,$$

donde τ_ε es tal, que $\mathbf{T}_{n_1 + n_2 - 2}(-\tau_\varepsilon, \tau_\varepsilon) = 1 - \varepsilon$ tendrá un nivel de significación exactamente igual a $1 - \varepsilon$ y el mismo podrá ser utilizado para cualesquiera valores (y no sólo grandes) de n_1, n_2. Este criterio, que se denomina criterio de Student, también posee ciertas propiedades de optimización exacta (y no sólo asintótica) (veáse [57]).

Ejemplo 4. Sea $X \in \Phi_{\alpha, \sigma_1^2}$, $Y \in \Phi_{\alpha, \sigma_2^2}$. La hipótesis $\{\sigma_1 = \sigma_2\}$ se verifica cuando se desconoce α. Procediendo del mismo modo que en el ejercicio anterior, llegaremos al valor R_1 en (8), cuyo denominador equivale al del ejemplo anterior, y el numerador es igual a

$$\sup_{\alpha, \sigma_1 \sigma_2} f_{(\alpha, \sigma_1^2)}(X) f_{(\alpha, \sigma_2^2)}(Y). \tag{15}$$

Escribiendo las ecuaciones para el punto del valor máximo, obtenemos

$$\sigma_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i - \alpha)^2 = S_X^2 + (\bar{x} - \alpha)^2,$$

$$\sigma_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(y_i - \alpha)^2 = S_Y^2 + (\bar{y} - \alpha)^2,$$

$$\frac{n_1}{\sigma_1^2}(\bar{x} - \alpha) + \frac{n_2}{\sigma_2^2}(\bar{y} - \alpha) = 0.$$

De aquí, poniendo

$$p = \frac{a}{\sigma_1^2} \cdot \frac{1}{a/\sigma_1^2 + (1 - a)/\sigma_2^2} \in (0,\ 1), \tag{16}$$

hallamos

$$\alpha = p\bar{x} + (1 - p)\bar{y},$$

$$\sigma_1^2 = S_X^2 + (1 - p)^2\Delta^2, \quad \sigma_2^2 = S_Y^2 + p^2\Delta^2,$$

donde, para abreviar, hemos supuesto que $\Delta = \bar{x} - \bar{y}$; p puede considerarse como la solución de la ecuación (16) o

$$p = \frac{a(S_Y^2 + p^2\Delta^2)}{a(S_Y^2 + p^2\Delta^2) + (1 - a)(S_X^2 + (1 - p)^2\Delta^2)}.$$

Como el máximo en (15) es igual a

$$(2\pi e)^{-(n_1+n_2)/2}(S_X^2 + (1 - p)^2\Delta^2)^{-n_1/2}(S_Y^2 + p^2\Delta^2)^{-n_2/2}, \tag{17}$$

comparándolo con (13) y (7), obtenemos el criterio asintóticamente minimax

$$\frac{aS_X^2 + (1 - a)S_Y^2 + a(1 - a)\Delta^2}{(S_X^2 + (1 - p)^2\Delta^2)^a(S_Y^2 + p^2\Delta^2)^{1-a}} > e^{h_\varepsilon(n_1+n_2)} \tag{18}$$

o bien

$$\frac{aS_X^2 + (1 - a)S_Y^2}{S_X^{2a}S_Y^{2(1-a)}} > e^{h_\varepsilon/(n_1+n_2)}A^{-1}, \tag{19}$$

donde $A = \dfrac{1 + \dfrac{a(1 - a)\Delta^2}{aS_X^2 + (1 - a)S_Y^2}}{(1 + (1 - p)^2\Delta^2/S_X^2)^a(1 + p^2\Delta^2/S_Y^2)^{1-a}}$, h_ε es una cuantila de la distribución χ^2 con un solo grado de libertad. Aquí

$$\Delta^2 = (\sigma_1^2/n_1 + \sigma_2^2(n_2)\xi^2, \ \xi \in \Phi_{0,1}, \ S_X^2/\sigma_1^2 \underset{P}{\to} 1, \ S_Y^2/\sigma_2^2 \underset{P}{\to} 1, \ \sigma_1^2/\sigma_2^2 \to 1,$$

$p \to a$ (para abreviar podemos considerar que $a = \frac{n_1}{n_1 + n_2}$ es fijo), $\ln A \underset{P}{\to} 0$ para cada una de las hipótesis semejantes que se examinan. Por consiguiente, el segundo miembro en (19) tiene la forma

$$1 + \frac{h_\varepsilon + \delta_n}{n_1 + n_2}, \qquad \delta_n \underset{P}{\to} 0.$$

El primer miembro de (19) es la relación entre la media aritmética y la media geométrica de los valores de S_X^2 y S_Y^2. Si se designa $S_X^2/S_Y^2 = Z^2$, la desigualdad inversa a (19) puede ser escrita en la forma

$$\frac{aZ^2 + (1-a)}{Z^{2a}} - 1 \leqslant \frac{h_\varepsilon + \delta_n}{n_1 + n_2}. \tag{20}$$

Aquí, en el primer miembro se halla la función de Z convexa hacia abajo (para evidenciar la exposición podemos considerar $a \leqslant 1/2$) que tiene un cero múltiplo en el punto $Z = 1$. Como el segundo miembro de esta desigualdad es pequeño, conviene hallar la solución en forma de $Z^2 = 1 + \zeta$ cuando ζ es pequeño. Utilizando el desarrollo en serie respecto a las potencias de ζ, y eliminando los términos del tercero y mayores órdenes de pequeñez, obtenemos, para las fronteras ζ_1, ζ_2 del intervalo donde es válida (20), los valores

$$\zeta_1 = \sqrt{\frac{2(h_\varepsilon + \delta_n')}{a(1-a)(n_1+n_2)}}, \quad \zeta_2 = \sqrt{\frac{2(h_\varepsilon + \delta_n'')}{a(1-a)(n_1+n_2)}},$$

$$\delta_n' \underset{P}{\to} 0, \quad \delta_n'' \underset{P}{\to} 0.$$

Esto significa que, si volvemos a las variables iniciales, el dominio

$$\sqrt{\frac{1}{2}a(1-a)(n_1+n_2)}\,|S_X^2/S_Y^2 - 1| > \sqrt{h_\varepsilon} = \lambda_{\varepsilon/2} \tag{21}$$

(λ_ε ha sido definido en el ejemplo 3) definirá el criterio asintóticamente equivalente a (18) y, por lo tanto, asintóticamente minimax.

Aquí al igual que en el ejemplo 3, podemos hacer que el criterio obtenido sea exacto, ya que conocemos la distribución precisa de la estadística S_X^2/S_Y^2. En efecto,

$$n_1 S_X^2/\sigma_1^2 \in \mathbf{H}_{n_1-1}, \; n_2 S_Y^2/\sigma_2^2$$

y en el caso de la hipótesis $H_1 = \{\sigma_1 = \sigma_2\}$,

$$\frac{n_1 S_X^2}{n_2 S_Y^2} \in \mathbf{F}_{n_1-1, n_2-1},$$

donde $\mathbf{F}_{n_1-1,n_2-1}$ es la distribución de Fisher introducida en el § 2.2 y tabulada en los manuales de estadística matemática. Esto significa que es posible calcular el nivel exacto de significación del criterio (21) y aplicarlo para cualesquiera n_1 y n_2 (las propiedades exactas de optimización de este criterio se exponen en [57]). Si son grandes los valores de n_1 y n_2, el primer miembro en (21) (sin signo de valor absoluto) es asintóticamente normal con parámetros (0, 1).

4. Criterio asintóticamente minimax para el problema de homogeneidad parcial. Supongamos que $X \in \mathbf{P}_{\theta_1}$, $Y \in \mathbf{P}_{\theta_2}$, $\theta_i = (u_i, v_i)$, $i = 1, 2$. Se verifica la hipótesis $\{u_1 = u_2\}$ frente a $\{u_1 \neq u_2\}$ cuando los valores de v_1 y v_2 en las muestras X e Y pueden ser cualesquiera. La dimensión u_i, al igual que antes, se designa por l, $l > k$.

Introduzcamos el nuevo parámetro $\bar{\theta} = (\theta_1, \theta_2) = (u_1, v_1, u_2, v_2)$ de dimensión $2k$. Al igual que antes, representemos la muestra (X, Y) (cuando $n_1 = n_2 = n$) como muestra con observaciones $(x_1, y_1), \ldots, (x_n, y_n)$ de densidad

$$f_{\bar{\theta}}(x, y) = f_{(u_1, v_1)}(x) f_{(u_2, v_2)}(y).$$

Para esta familia, el problema de homogeneidad parcial equivale al problema de verificación de la hipótesis H_1, el cual consiste en que $\bar{\theta}$ permanece en la "curva" $\theta = g(\alpha) = (u_1, v_1, u_1, v_2)$, donde $\alpha = (u_1, v_1, v_2)$ es el "subparámetro" de dimensión $2k - l$. Le proponemos al lector que escriba, siguiendo los razonamientos de los dos apartados anteriores, la matriz $G = \left\| \dfrac{\partial g_i}{\partial \alpha_j} \right\|$, $i = 1, \ldots, 2k$, $j = 1, \ldots, 2k - l$. Su rango es igual a $2k - l$.

Al igual que en los apartados 2 y 3, consideraremos "localizado" el problema cerca del punto $\theta_0 = (u_0, v_0)$. A la par con θ introduzcamos el parámetro $\tau = \tau(\bar{\theta}) = (\tau', \tau'', \tau''', \tau^{IV}) = (u_1 - u_0, v_1 - v_0, u_2 - u_1, v_2 - v_0)$. La transformación inversa $\bar{\theta} = \bar{\theta}_{(\tau)}$ tiene las coordenadas

$$u_1 = \tau' + u_0, \; v_1 = \tau'' + v_0, \; u_2 = \tau''' + u_0,$$
$$v_2 = \tau^{IV} + v_0.$$

Si se pone $\tau = \gamma/\sqrt{n}$, $\gamma = (\gamma', \gamma'', \gamma''', \gamma^{IV})$, la hipótesis H_1 tendrá la forma $H_1 = \{\gamma''' = 0\}$. En calidad de alternativa consideraremos la hipótesis "aislada" $H_2^b = \{\gamma''' I_1(\theta_0) \gamma'''^T \geqslant b^2\}$, donde $I_1(\theta)$ tiene el mismo sentido que en el teorema 2.

Teorema 3. *Supongamos que en el entorno del punto θ_0, la familia $\{\mathbf{P}_\theta\}$ satisface las condiciones (RR). Entonces, el criterio de la relación de verosimilitud*

$$R_1(X, Y) = \frac{\sup\limits_{(\theta_1, \theta_2)} f_{\theta_1}(X) f_{\theta_2}(Y)}{\sup\limits_{(u, v_1, v_2)} f_{(u, v_1)}(X) f_{(u, v_2)}(Y)} > e^{h_\varepsilon/2} \tag{22}$$

es el criterio asintóticamente minimax de nivel asintótico $1 - \varepsilon$ *para verificar* H_1 *frente a la hipótesis* H_1^{ε} *definida en* (9), *para los valores arbitrarios de* v_1 *y* v_2. *El valor de* h_ε *aquí es el mismo que en el teorema* 2.

La demostración de este teorema repite los razonamientos de los apartados precedentes y asimismo se basa por completo en el teorema 3.15.4. Le dejamos al lector que él mismo determine la matriz de información de Fisher $\bar{I}(\bar{\theta})$ para el parámetro $\bar{\theta}$, y la matriz M_2 para la familia de densidad $f_{\bar{\theta}(0,\,0,\,\beta,\,0)} = f_{(u_0,\,v_0,\,u_0+\beta,\,v_0)}$ en el punto $\beta = 0$.

Con ayuda de la matriz $\bar{I}((\hat{\theta}_X^*, \hat{\theta}_Y^*))$ y los vectores $(\hat{\theta}_X^*, \hat{\theta}_Y^*) - (u^*, v_1^*, u^*, v_2^*)$, donde $(\hat{\theta}_X^*, \hat{\theta}_Y^*)$ y (u^*, v_1^*, v_2^*) son los vectores en los que se alcanzan los valores máximos del numerador y el denominador en (22), es posible, como antes, mediante el teorema 3.15.4 (véase (3.15.12), construir el criterio asintóticamente equivalente que utiliza la forma cuadrática de las estimaciones introducidas. ◁

Ejemplo 5. *Comparación de las varianzas de las poblaciones normales. Sea* $X \in \Phi_{\alpha_1,\,\sigma_1^2}$, $Y \in \Phi_{(\alpha_2,\,\sigma_2^2)}$, $H_1 = \{\sigma_1 = \sigma_2\}$. Aquí, los cálculos son mucho más fáciles que en el ejemplo 4, ya que conocemos el valor del numerador en (22) (al igual que el vector $(\hat{\theta}_X^*, \hat{\theta}_Y^*) = (\bar{x}, S_X^2, \bar{y}, S_Y^2)$), y el valor del denominador ha sido hallado en el ejemplo 3 (véase (12)). La desigualdad (22) aquí tendrá la forma

$$\frac{aS_X^2 + (1-a)S_Y^2}{S_X^{2a} S_Y^{2(1-a)}} > e^{h_\varepsilon/(n_1+n_2)}.$$

Comparando esto con (19) y con los planteamientos posteriores, llegaremos a los mismos criterios y a las mismas deducciones que en el ejemplo 4.

Ejemplo 6. *Problema de Behrens — Fisher acerca de la comparación de las medias de dos poblaciones normales.* Sea $X \in \Phi_{\alpha_1,\,\sigma_1^2}$, $Y \in \Phi_{\alpha_2,\,\sigma_2^2}$, $H_1 = \{\alpha_1 = \alpha_2\}$ y supongamos que los valores σ_1 y σ_2 son arbitrarios. Para este ejemplo, el numerador en (22) es el mismo que en el párrafo anterior, y el denominador fue hallado en el ejemplo 4 (véase (17); allí éste era el numerador para (8)).

Por consiguiente, el criterio asintóticamente minimax tiene la forma

$$\left(\frac{S_X^2 + (1-p)^2\Delta^2}{S_X^2}\right)^a \left(\frac{S_Y^2 + p^2\Delta^2}{S_Y^2}\right)^{1-a} > e^{h_\varepsilon/(n_1+n_2)}; \tag{23}$$

aquí $\Delta = \bar{x} - \bar{y}$ es representable en la forma

$$\Delta = (\alpha_1 - \alpha_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\,\xi, \quad \xi \in \Phi_{0,1}.$$

$$S_X^2/\sigma_1^2 \xrightarrow[p]{} 1, \quad S_Y^2/\sigma_2^2 \xrightarrow[p]{} 1,$$

así que $\Delta \underset{P}{\to} 0$ para la hipótesis H_1. Esta relación, evidentemente, también conserva su validez para cada una de las alternativas semejantes. Para hallar un criterio más simple en cuanto a su forma y que equivalga asintóticamente a (23), en ambos miembros de la desigualdad (23) separaremos sus partes principales. Obtendremos

$$\frac{a(1-p)^2\Delta^2}{S_X^2} + \frac{(1-a)p^2\Delta^2}{S_Y^2} + \Delta^4\varrho_n > \frac{h_\varepsilon}{n_1+n_2} + O\left(\frac{1}{(n_1+n_2)^2}\right),$$

donde $\varrho_n \underset{P}{\to} \varrho = \text{const}$. Teniendo en cuenta que

$$p = \frac{aS_Y^2}{aS_Y^2 + (1-a)S_X^2} + \Delta^2\varrho_n',$$

$\varrho_n' \underset{P}{\to} \varrho' = \text{const}$, obtenemos

$$\frac{a(1-a)^2 S_X^2\Delta^2(n_1+n_2) + a^2(1-a)S_Y^2\Delta^2(n_1+n_2)}{(aS_Y^2 + (1-a)S_X^2)^2} +$$

$$+ \Delta^4(n_1+n_2)\varrho_n'' > h_\varepsilon + O\left(\frac{1}{n_1+n_2}\right),$$

donde $\varrho_n'' \to \varrho'' = \text{const}$, $\Delta^4(n_1+n_2) \underset{P}{\to} 0$. Equivalentemente esto se puede escribir de la forma siguiente:

$$\frac{\Delta^2(n_1+n_2)}{S_X^2/a + S_Y^2/(1-a)} > h_\varepsilon + \delta_n, \quad \delta_n \underset{P}{\to} 0.$$

De aquí se deduce que el criterio

$$\frac{|\bar{x} - \bar{y}|\sqrt{n_1+n_2}}{\sqrt{S_X^2/a + S_Y^2/(1-a)}} > \sqrt{h_\varepsilon} = \lambda_{\varepsilon/2} \tag{24}$$

es asintóticamente equivalente a (23) y, por lo tanto, asintóticamente minimax para el problema de Behrens — Fisher. Aquí $\lambda_{\varepsilon/2}$ tiene el mismo sentido que en el ejemplo 4. A distinción de los ejemplos 2—4, aquí la distribución antelímite de la estadística en el primer miembro (24) depende, para la hipótesis H_1, de los parámetros σ_1^2 y σ_2^2.

5. Algunos otros problemas. Aquí señalaremos dos clases más de problemas cuya solución asintótica puede ser hallada con ayuda del teorema 3.15.4.

1) A la primera clase de problemas pertenecen aquéllos que generalizan los problemas de los apartados 2—4 para el caso cuando se verifican las hipótesis de tipo $\{\theta_1 = f(\theta_2)\}$ (por ejemplo, $\{\theta_1 = a + b\theta_2\}$) en condiciones del apartado 2, y de tipo $\{u_1 = f(u_2)\}$ en condiciones de los apartados 3 y 4. Es fácil notar que los planteamientos de los apartados 2—4 se extienden a este caso más general.

2) A la segunda clase de problemas pertenecen aquéllos que constan de tres muestras y más. Examinemos, por ejemplo, el problema de homogeneidad para tres muestras. Supongamos que $X \in \mathbf{P}_{\theta_1}$, $Y \in \mathbf{P}_{\theta_2}$, $Z \in \mathbf{P}_{\theta_3}$. Se verifica la hipótesis $H_1 = \{\theta_1 = \theta_2 = \theta_3\}$ frente a la alternativa adicional. Supongamos, para abreviar, que los volúmenes n_1, n_2 y n_3 de las muestras son iguales a $n_1 = n_2 = n_3 = n$. Examinemos la muestra unida (X, Y, Z) como una muestra de volumen n con observaciones (x_1, y_1, z_1), ..., (x_n, y_n, z_n) de densidad $f_{\bar\theta}(x, y, z) = f_{\theta_1}(x)f_{\theta_2}(y)f_{\theta_3}(z)$, donde $\bar\theta = (\theta_1, \theta_2, \theta_3)$. Entonces, la hipótesis H_1 será equivalente al hecho de que $\bar\theta$ permanece en la "curva" $\bar\theta = g(\alpha)$, $\alpha \equiv \theta_1$, $g(\alpha) = (\alpha, \alpha, \alpha)$. Vemos que el problema de nuevo se reduce al problema examinado en el teorema 3.15.4.

§ 2. Problema de homogeneidad en el caso general

1. Planteamiento del problema. En este párrafo examinaremos dos muestras X e Y de volúmenes n_1 y n_2, respectivamente, sin suponer que las mismas pertenecen a cualquier familia paramétrica.

El problema de homogeneidad de las muestras X e Y, en el caso general consiste en lo siguiente. Designemos por $\mathbf{P}_1$ y $\mathbf{P}_2$ las distribuciones de las muestras X e Y: $X \in \mathbf{P}_1$, $Y \in \mathbf{P}_2$. Se verifica la hipótesis $H_1 = \{\mathbf{P}_1 = \mathbf{P}_2\}$ frente a $H_2 = \{\mathbf{P}_1 \neq \mathbf{P}_2\}$. Evidentemente, ambas hipótesis son compuestas. Las distribuciones $\mathbf{P}_1$ y $\mathbf{P}_2$ pueden elegirse de una familia dada $\mathscr{P}$ o ser arbitrarias. El principio general de construcción del criterio estadístico para verificar H_1 frente a H_2 es el mismo que en el capítulo 3. Al igual que en el § 1, la diferencia sólo consiste en que aquí este principio se basa en la muestra unida (X, Y), así que $\pi = \pi(X, Y)$ es la probabilidad de aceptar H_2 para una muestra dada (X, Y). En el caso no randomizado ($\pi = 0$ ó 1), el criterio π es definido por una región crítica $\Omega \subset \mathscr{X}^{n_1+n_2}$ tal, que para $(X, Y) \in \Omega$ se acepta H_2. El número

$$1 - \varepsilon = \inf_{\mathbf{P}_1 \in \mathscr{P}} \mathbf{P}_1 \times \mathbf{P}_1((X, Y) \notin \Omega)$$

se llama *nivel de significación*, y el valor

$$\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2((X, Y) \in \Omega), \quad \mathbf{P}_1 \in \mathscr{P}, \quad \mathbf{P}_2 \in \mathscr{P}$$

se denomina potencia del criterio π en el "punto" $(\mathbf{P}_1, \mathbf{P}_2)$.

El criterio π se denomina criterio *conciliable* si $\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) \to 1$ cuando $n_1 \to \infty$, $n_2 \to \infty$ y para todas $\mathbf{P}_1 \neq \mathbf{P}_2$, $\mathbf{P}_1 \in \mathscr{P}$, $\mathbf{P}_2 \in \mathscr{P}$.

Ya sabemos que con el crecimiento de n_1 y n_2, las distribuciones empíricas $\mathbf{P}^*_X$, $\mathbf{P}^*_Y$, correspondientes a las muestras X e Y, se aproximan indefinidamente a $\mathbf{P}_1$ y $\mathbf{P}_2$, respectivamente. Por eso, la base natural para construir los criterios de homogeneidad es el uso de distintos tipos de "distancias" $d(\mathbf{P}^*_X, \mathbf{P}^*_Y)$ entre $\mathbf{P}^*_X$ y $\mathbf{P}^*_Y$, donde d satisface las mismas condiciones genera-

les que hemos descrito en el § 3.12. En este caso revisten interés especial los criterios no paramétricos y asintóticamente no paramétricos que se definen del modo siguiente.

Sea $d(\mathbf{P}, \mathbf{Q})$ cierta distancia (no obligatoriamente métrica) en el espacio de distribuciones. Si la probabilidad

$$\mathbf{P}_1 \times \mathbf{P}_1(d(\mathbf{P}_X^*, \mathbf{P}_Y^*) > c) = \varepsilon \tag{1}$$

no depende de la muestra $\mathbf{P}_1$, entonces el criterio π, definido por las igualdades

$$\pi(X, Y) = \begin{cases} 0, \text{ si } d(\mathbf{P}_X^*, \mathbf{P}_Y^*) \leqslant c, \\ 1, \text{ en el caso contrario,} \end{cases} \tag{2}$$

se llama criterio *no paramétrico*. Es evidente que el criterio no paramétrico construido tendrá un igual a nivel $1 - \varepsilon$.

Así mismo se determinan los criterios no paramétricos cuando (1) se conserva asintóticamente al introducir la operación $\lim\limits_{n_1 \to \infty,\ n_2 \to \infty}$ en el primer miembro. En este caso el criterio (2) tendrá un nivel asintótico igual $1 - \varepsilon$. Cuando falta la no parametricidad (exacta o asintótica) es muy difícil construir los criterios de verificación de la homogeneidad de un nivel dado.

Examinemos algunos criterios principales de verificación de la homogeneidad.

2. Criterio de Kolmogórov — Smirnov. Supongamos que $\mathbf{P}_1$ y $\mathbf{P}_2$ pertenecen a la clase $\mathscr{P}$ de todas las distribuciones continuas en una recta, y que F_X^* y F_Y^* son funciones empíricas de distribución, correspondientes a $\mathbf{P}_X^*$ y $\mathbf{P}_Y^*$. En calidad de distancia $d(\mathbf{P}_X^*, \mathbf{P}_Y^*)$, el criterio de Kolmogórov — Smirnov considera la estadística

$$D_{n_1, n_2} = \sup_t |F_X^*(t) - F_Y^*(t)|.$$

El criterio $D_{n_1, n_2} > c$, construido con ayuda de la estadística D_{n_1, n_2} no es paramétrico. En efecto, supongamos que es cierta la hipótesis H_1 y que $F(t)$ es la función general de distribución de X e Y. La estadística D_{n_1, n_2} se puede escribir de la forma siguiente:

$$D_{n_1 n_2} = \sup_t |G_X^*(F(t)) - G_Y^*(F(t))|, \tag{3}$$

donde $G_X^*(u) = F_X^*(F^{-1}(u))$ es la función empírica de distribución que corresponde a la distribución uniforme en [0, 1] (veánse los §§ 1.6 y 3.12). Pero en virtud de (3), $D_{n_1, n_2} = \sup\limits_u |G_X^*(u) - G_Y^*(u)|$, así que la distribución D_{n_1, n_2} no depende de F de ningún modo.

Se puede hallar la distribución exacta de la estadística D_{n_1, n_2}. Por ejemplo, cuando $n_1 = n_2 = n$,

$$\mathbf{P}(nD_{n,n} \geqslant k) = 2(C_{2n}^n)^{-1} \sum_{j=1}^{[n/k]} (-j)^{j+1} C_{2n}^{n-jk}, \tag{4}$$

$k = 1, 2, \ldots, n$. Este hecho fue establecido por Gnedendo y Koroliuk reduciendo esta tarea al simple problema de vagancias aleatorias (véase [32]).

En el § 1.6 hemos visto que la distribución $n_1 G^*_X(u)$ coincide con la distribución del proceso poissoniano $\zeta_1(u)$ a condición de que $\zeta_1(1) = n_1$. Como $G^*_X(u)$ y $G^*_Y(y)$ son independientes, la distribución $G^*_X(u) - G^*_Y(u)$, $u \in [0, 1]$ coincide con la distribución del proceso poissoniano compuesto $\zeta(u)$, en el que, con intensidad n_1, se producen saltos de magnitud $1/n_1$, y con intensidad n_2, saltos de magnitud $1/n_2$; la distribución ha de tomarse a condición de que ocurrieron $n_1 + n_2$ saltos y que $\zeta(1) = 0$. Por eso

$$\mathbf{P}(D_{n_1,\, n_2} < x) = \mathbf{P}(\sup_{u \leqslant 1} |\zeta(u)| < x/\zeta(1) = 0;\ \textit{ocurrieron } n_1 + n_2 \textit{ saltos}).$$

A base de este hecho, en el Suplemento II, además del teorema 1.6.2 de convergencia del proceso $w_n(u) = \sqrt{n_1}(G^*_X(u) - u)$ hacia el puente browniano $w^\circ(u)$, también se demuestra la afirmación de que hacia el referido puente también converge el proceso

$$w_{n_1,\, n_2}(u) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,(G^*_X(u) - G^*_Y(u)).$$

Mejor dicho, para cualquier funcional f medible y continua en una métrica uniforme, la distribución $f(w_{n_1,\, n_2})$ converge hacia la distribución $f(w^\circ)$. De aquí se deduce inmediatamente la siguiente afirmación denominada teorema de Smirnov.

Teorema 1.

$$\lim_{n_1 \to \infty,\ n_2 \to \infty} \mathbf{P}\left(\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\, D_{n_1,\, n_2} < x\right) = \mathbf{P}\left(\sup_{u \leqslant 1} |w^\circ(u)| < x\right) = K(x),$$

donde $K(x)$ es la función de Kolmogórov (véanse los §§ 1.8 y 3.12).

Como la función $K(x)$ está tabulada, el teorema 1 ofrece un medio cómodo para el cálculo aproximado del nivel de significación del criterio de Kolmogórov — Smirnov.

Le dejamos al lector que el mismo se cerciore de que el criterio de Kolmogórov — Smirnov es conciliable.

3. Criterio de signos. Sea $n_1 = n_2 = n$. Entonces, de las observaciones de las muestras X e Y se pueden componer n diferencias:

$$\mathrm{x}_1 - \mathrm{y}_1, \ldots, \mathrm{x}_n - \mathrm{y}_n. \tag{5}$$

Si es cierta la hipótesis H_1 y $\mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 = 0) = 0$ para todas las $\mathbf{P}_1 \in \mathscr{P}$ (esto, evidentemente, siempre es así cuando $\mathscr{P}$ es un conjunto de distribuciones continuas), entonces

$$\mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 > 0) = \mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 < 0) = 1/2.$$

La estadística ν del criterio de signos es el número de diferencias positivas en (5)[*]. El propio criterio se puede construir adoptando en calidad de conjunto crítico,

$$\Omega = \left\{(X, Y): \left|\nu - \frac{n}{2}\right| > c\right\}.$$

Como la distribución de ν no depende de $\mathbf{P}_1$,

$$\mathbf{P}_1 \times \mathbf{P}_1(\nu = k) = C_n^k 2^{-n},$$

por lo tanto, este criterio no es paramétrico.

El número c, según el nivel dado $1 - \varepsilon$ del criterio, se elige de la relación

$$k:|2k - n| \leqslant 2c\, C_n^k 2^{-n} \geqslant 1 - \varepsilon. \tag{6}$$

Como aquí el primer miembro crece de un modo discreto con el aumento de c, en calidad de solución conviene tomar el valor mínimo de c, con el que el primer miembro en (6) supera el valor de $1 - \varepsilon$.

Vemos que aquí se utiliza el criterio para verificar la hipótesis de que la probabilidad de éxito en el esquema de Bernoulli es igual a 1/2. Desde el punto de vista del problema inicial, se verifica no la hipótesis de homogeneidad, sino una hipótesis más amplia acerca de que

$$\mathbf{P}_1 \times \mathbf{P}_2(x_1 - y_1 < 0) = \int F_1(t)dF_2(t) = 1/2, \tag{7}$$

donde F_i corresponde a $\mathbf{P}_i$, $i = 1, 2$. La relación (7) significa que la mediana de distribución $x_1 - y_1$ es igual a 0.

El criterio de los signos del nivel asintótico $1 - \varepsilon$ tendrá la forma siguiente:

$$\pi(X, Y) = 1, \quad si \quad \frac{2\left|\nu - \frac{n}{2}\right|}{\sqrt{n}} > \lambda_{\varepsilon/2}, \tag{8}$$

$$\Phi_{0,1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) = 1 - \varepsilon.$$

Este criterio no es conciliable, ya que para $\mathbf{P}_1 \neq \mathbf{P}_2$ que satisfacen (7), $\beta_n(\mathbf{P}_1, \mathbf{P}_2) \to \varepsilon < 1$ cuando $n_1 \to \infty$, $n_2 \to \infty$.

4. Criterio de Wilkoxon. Este criterio se aplica ampliamente al verificar las hipótesis de homogeneidad.

Juntemos las muestras X e Y en una sola muestra (X, Y) y construyamos de ella una serie variacional, o sea, situemos todas las observaciones

[*] Si en las muestras X e Y, debido al valor aproximado de los datos, resulta que algunas diferencias $x_i - y_i = 0$, entonces, éstas deben ser simplemente omitidas, tomando en calidad de n el número de diferencias distintas del cero.

en orden de crecimiento. Obtendremos una sucesión de tipo

$$y^{(1)},\ y^{(2)},\ x^{(3)},\ y^{(4)},\ x^{(5)},\ \ldots, \tag{9}$$

donde el índice superior designa el número de observación en la serie variacional general, mientras que la letra indica la pertenencia a la muestra. Supongamos que $r_1, r_2, \ldots, r_n$, designan los números de elementos de la muestra X en la serie variacional (9). Para la sucesión escrita en (9), $r_1 = 3$, $r_2 = 5$. Llámase *estadística de Wilkoxon* la función

$$U = U(X, Y) = \sum_{i=1}^{n} (r_i - i),$$

donde $r_i - i$ es el número de elementos de la muestra Y que son menores de $x_{(i)}$.

En vista de que el orden de observaciones en (9) es invariante respecto a las transformaciones monótonas de las variables (el orden de $F_X^*(t)$, $F_Y^*(t)$ será el mismo que para $F_X^*(F^{-1}(t))$, $F_Y^*(F^{-1}(t))$, donde F es la función de distribución), el criterio construido según la estadística U no será paramétrico.

Teorema 2. *Supongamos que $X \in \mathbf{P}_1$, $Y \in \mathbf{P}_2$ y $F_i \in \mathcal{F}$ son las funciones de distribución correspondientes a $\mathbf{P}_i$, $i = 1, 2$; $\mathcal{F}$ es la clase de todas las funciones de distribución continuas. Supongamos también, que $a = n_1/(n_1 + n_2) \to a_0$ cuando $n_1 \to \infty$, $n_2 \to \infty$. Entonces*

$$\frac{U - n_1 n_2 \mathbf{M} F_2(x_1)}{\sqrt{n_1 n_2 (n_1 + n_2)}} \in \Phi_{0,\sigma^2}, \tag{10}$$

donde $\sigma^2 = (1 - a_0)\mathbf{D}F_2(x_1) + a_0 \mathbf{D}F_1(y_1)$.

Si $F_1 = F_2 = F$, entonces $F_2(x_1) \in \mathrm{U}_{0,1}$, $F_1(y_1) \in \mathrm{U}_{0,1}u$, por consiguiente, $\mathbf{M}F_2(x_1) = 1/2$, $\mathbf{D}F_2(x_1) = \mathbf{D}F_1(y_1) = 1/12$.

Por lo tanto, el criterio de Wilkoxon de nivel asintótico $1 - \varepsilon$ tendrá la forma siguiente:

$$\left| U - \frac{n_1 n_2}{2} \right| > \frac{\lambda_{\varepsilon/2}\sqrt{n_1 n_2 (n_1 + n_2)}}{2\sqrt{3}}, \tag{11}$$

$$\Phi_{0,1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) = 1 - \varepsilon.$$

De (10) se deduce que este criterio tiene por objeto principal la verificación de la hipótesis (compárese con (7))

$$\int F_2(t)dF_1(t) = 1/2 \text{ o bien } \int (F_2(t) - F_1(t)dF_1(t) = 0. \tag{12}$$

Si admitimos, sin limitar la generalidad, que $F_1(t) \equiv t$, $t \in [0, 1]$, y si suponemos que $F_2(0) = 0$, $F_2(1) = 1$, entonces, en virtud de la igualdad

$$\int_0^1 (1 - F_2(t))dt = \mathbf{M}y_1,$$

la hipótesis que se verifica adoptará la forma $y_1 = 1/2$.

Esto significa que el criterio de Wilkoxon, al igual que el criterio de signos, es principalmente sensible a los desplazamientos de las distribuciones una respecto a otra. Para tales alternativas desplazadas, su potencia puede ser bastante grande (véase el ejemplo 1). Pero si $F_2 \not\equiv F_1$ y se cumple (12), entonces, según el criterio de Wilkoxon, la hipótesis $\{F_2 = F_1\}$ será aceptada con una probabilidad próxima a $\Phi_{0,1}\left(-\frac{\lambda_{\varepsilon/2}}{2\sqrt{3}\sigma}, \frac{\lambda_{\varepsilon/2}}{2\sqrt{3}\sigma}\right)$. Esto significa que el criterio de Wilkoxon será inconciliable.

Demostración del teorema 2. La estadística U puede ser escrita de la forma siguiente:

$$U = \sum_{i=1}^{n_1} n_2 F_Y^*(x_i) = n_1 n_2 \int F_Y^*(t) dF_X^*(t).$$

Designemos

$$w_X(t) = \sqrt{n_1}(F_X^*(t) - F_1(t)), \quad w_Y(t) = \sqrt{n_2}(F_Y^*(t) - F_2(t)).$$

Entonces es evidente que

$$U = n_1 n_2 \int F_2(t) dF_1(t) + \sqrt{n_1 n_2 (n_1 + n_2)} \times$$
$$\times \left[\sqrt{a} \int w_Y(t) dF_1(t) + \sqrt{1 - a} \int F_2(t) dw_X(t)\right] + \sqrt{n_1 n_2} \int w_Y(t) dw_X(t). \tag{13}$$

Como aquí $\int F_2)t)dw_X(t) = \int w_X(t)dF_2(t)$ y, por consiguiente, las integrales segunda y tercera en (13) tienen la misma forma y son independientes, para demostrar el teorema es suficiente convencerse de que

$$\int w_Y(t) dF_1(t) \in \Phi_{0,\sigma_1^2}, \quad \sigma_1^2 = \mathbf{D}F_1(y_1) \tag{14}$$

y que

$$\frac{1}{\sqrt{n_1 + n_2}} \int w_Y(t) dw_X(t) \underset{P}{\to} 0. \tag{15}$$

En virtud del teorema 1.6.2,

$$\int w_Y(t) dF_1(t) \in \int w^\circ(F_2(t)) dF_1(t), \tag{16}$$

donde $w^\circ(u)$ es el puente borwniano. Para hallar la distribución de la última integral, señalaremos que las trayectorias del proceso wieneriano $w(u)$ de probabilidad 1 son continuas [11], $w^\circ(u) = w(u) - uw(1)$, y que, por lo tanto, la integral (16) es, por definición, el resultado de la convergencia casi segura de las sumas cuando $N \to \infty$,

$$\sum_{i=1}^{N} w(F_2(t_i))\Delta_i F_1 - m_1 w(1), \tag{17}$$

donde $m_1 = \int F_2(t)dF_1(t)$, $\{t_i\}_{i=0}^{N}$ forman la partición del eje real, $\Delta_i g = g(t_i) - g(t_{i-1})$, $w(F_2(t_i)) = \sum_{l=1}^{i} \Delta_l w(F_2)$, $w(1) = \sum_{l=1}^{N} \Delta_l w(F_2)$.

En virtud de la transformación de Abel,

$$\sum_{i=1}^{N}\left(\sum_{l=1}^{i} d_l\right)b_i = \sum_{l=1}^{N}\left(\sum_{i=l}^{N} b_i\right)d_l.$$

Por eso (17) es igual a

$$\sum_{l=1}^{N} (1 - F_1(t_{l-1}) - m_1)\Delta_l w(F_2). \tag{18}$$

Aquí $1 - m_1 = \int F_1(t)dF_2(t) = m_2$ y $\Delta_l w(F_2)$ son variables aleatorias normalmente distribuidas e independientes con parámetros $(0, \Delta_l F_2)$. Por eso la distribución (17), (18) será normal con media nula y con varianza

$$\sum_{l=1}^{N} (m_2 - F_1(t_{l-1}))^2\Delta_l F_2 \to \int (m_2 - F_1(t))^2 dF_2(t) = \mathbf{D}F_1(y_1).$$

La relación (14) queda demostrada.

Para demostrar (15)*), lo más fácil es estimar la varianza de la integral en (15). Volviendo a aproximar la integral con ayuda de la suma final, es posible convencerse que la varianza

$$D_{X,Y} = \mathbf{M}\left(\int w_Y(t)dw_X(t)\right)^2$$

está limitada cuando $n_1 \to \infty$, $n_2 \to \infty$. De aquí y de la desigualdad de Chébishev resulta (15). Debido a los cálculos voluminosos y rutinarios, omitiremos la demostración del carácter limitado de $D_{X,Y}$. ◁

Datos más exactos acerca de los criterios de signos y de Wilkoxon se exponen en [41].

Ejemplo 1. Hemos señalado que los criterios de signos y de Wilkoxon son los más sensibles a los desplazamientos. Por eso es interesante comparar su potencia con la del criterio óptimo en el problema donde la homogeneidad se verifica para la familia $\mathscr{P}$ de distribuciones que sólo se distinguen por sus desplazamientos. Pues, supongamos que

$$\mathscr{P} = \{\Phi_{\alpha,1}\},\quad \mathbf{P}_1 = \Phi_{\alpha_1,1},\quad \mathbf{P}_2 = \Phi_{\alpha_2,1},\quad n_1 = n_2 = n.$$

En este caso, conforme al teorema 1.1, para verificar la hipótesis $H_1 = \{\mathbf{P}_1 = \mathbf{P}_2\} = \{\alpha_1 = \alpha_2\}$ frente a $H_2^b = \{|\alpha_1 - \alpha_2| \geqslant b/\sqrt{n}\}$ existe el criterio asintóticamente minimax π_0 de nivel $1 - \varepsilon$, que tiene la forma

$$|\bar{x} - \bar{y}| > \lambda_{\varepsilon/2}\sqrt{2/n},\quad \Phi_{0,1}((-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2})) = 1 - \varepsilon$$

(el hecho de que en nuestro ejemplo esta desigualdad equivale a (1.3 y 1.4), el lector puede comprobarlo personalmente). Tomemos este criterio por

*) La integral en (15) converge respecto a la distribución hacia $\int w^0(F_2(t))dw^0(F_1(t))$.

patrón para la comparación con otros criterios y examinemos la alternativa $(\mathbf{P}_1, \mathbf{P}_2)$, donde $\alpha_2 = \alpha_1 + c/\sqrt{n}$ (examinamos las alternativas semejantes para no tratar el problema de grandes desviaciones). Es evidente que en este caso $(\bar{x} - \bar{y}) \in \Phi_{-c/\sqrt{n},\, 2/n}$. Por lo tanto,

$$\beta_{\pi_0}(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2(|\bar{x} - \bar{y}| > \lambda_{\varepsilon/2}\sqrt{2n}) =$$
$$= 1 - \Phi_{-c/\sqrt{2},1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) =$$
$$= 1 - \Phi_{0,1}(-\lambda_{\varepsilon/2} + c/\sqrt{2}, \lambda_{\varepsilon/2} + c/\sqrt{2}) \equiv \beta_0(c). \qquad (19)$$

Examinemos ahora el criterio de signos (8), designándolo por π_1. Haciendo uso del desarrollo en serie de las potencias de $c/\sqrt{n}$, hallamos $(\Phi_{\alpha,\sigma^2}(x) = \Phi_{\alpha,\sigma^2}((-\infty, x)))$

$$\mathbf{P}_1 \times \mathbf{P}_2(x_1 - y_1 < 0) = \Phi_{0,2}\left(\frac{c}{\sqrt{n}}\right) = \frac{1}{2} + \frac{c}{\sqrt{n}}\cdot\frac{1}{2\sqrt{\pi}} + O\left(\frac{1}{n}\right).$$

Por eso en el punto $(\mathbf{P}_1, \mathbf{P}_2)$

$$\frac{2}{\sqrt{n}}\left(\nu - \frac{n}{2} + \frac{c\sqrt{n}}{2\sqrt{\pi}}\right) \in \Phi_{0,1}.$$

Por consiguiente, para el criterio de signos π_1 de nivel asintótico $1 - \varepsilon$,

$$\beta\pi_1(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2\left(2\left|\nu - \frac{n}{2}\right| > \lambda_{\varepsilon/2}\sqrt{n}\right) \to$$
$$\to 1 - \Phi_{0,1}\left(\left(-\lambda_{\varepsilon/2} + \frac{c}{\sqrt{\pi}}, \lambda_{\varepsilon/2} + \frac{c}{\sqrt{\pi}}\right)\right).$$

Volvamos, por último, al criterio π_2 de Wilkoxon (véase (11)) que en nuestro caso tiene la forma

$$\left|U - \frac{n^2}{2}\right| > \frac{\lambda_{\varepsilon/2}n^{3/2}}{\sqrt{6}}.$$

Evidentemente, la estadística U es invariante respecto a la transformación de desplazamiento de los elementos de las muestras X e Y. Por eso se puede considerar que $\mathbf{P}_1 = \Phi_{0,1}$, $\mathbf{P}_2 = \Phi_{c/\sqrt{n},1}$ y, por lo tanto,

$$\mathbf{M}F_2(x_1) = \int F_2(t)dF_1(t) = \int \Phi_{0,1}\left(t - \frac{c}{\sqrt{n}}\right)d\Phi_{0,1}(t) =$$
$$= \Phi_{0,2}\left(-\frac{c}{\sqrt{n}}\right) = \frac{1}{2} - \frac{c}{\sqrt{n}}\cdot\frac{1}{2\sqrt{\pi}} + O\left(\frac{1}{n}\right).$$

Como $\mathbf{D}F_2(x_1) \to \mathbf{D}F_1(x_1) = 1/12$, $\mathbf{D}F_1(y_1) \to \mathbf{D}F_1(x_1) = 1/12$, según el teorema 2,

$$\beta_{\pi_2}(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2\left(\left|U - \frac{n^2}{2}\right| > \frac{\lambda_{\varepsilon/2} n^{3/2}}{\sqrt{6}}\right) =$$

$$= 1 - \mathbf{P}_1 \times \mathbf{P}_2\left(-\lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}} \leqslant \right.$$

$$\left. \leqslant \sqrt{6}n^{-3/2}\left(U - \frac{n^2}{2} + \frac{n^{3/2}c}{2\sqrt{\pi}}\right) \leqslant \lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}}\right) =$$

$$= 1 - \Phi_{0,1}\left(-\lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}},\ \lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}}\right).$$

Ahora debemos señalar que $\beta_0(c)$ (véase (19)) es una función monótona creciente de c y que, con grandes valores de n,

$$\beta_{\pi_1}(\mathbf{P}_1, \mathbf{P}_2) \approx \beta_0\left(\sqrt{\frac{2}{\pi}}\, c\right), \quad \beta_{\pi_2}(\mathbf{P}_1, \mathbf{P}_2) \approx \beta_0\left(\sqrt{\frac{3}{\pi}}\, c\right).$$

Ahora bien, para cada $c > 0$, el más potente entre los π_0, π_1 y π_2 resulta, como era de esperar, el criterio π_0. Le siguen el criterio de Wilkoxon y el de signos; con la particularidad de que el criterio de Wilkoxon cede muy poco al criterio π_0, ya que $\sqrt{3/\pi} \approx 0{,}977$.

Si para ese mismo desplazamiento $\alpha_2 - \alpha_1 = c/\sqrt{n}$ examinamos las muestras X' e Y' de nivel $n' > n$, entonces, para obtener (con ayuda de los cálculos efectuados) la potencia de los criterios $\pi_j(X', Y')$ en el punto $(\mathbf{P}_1, \mathbf{P}_2)$, debemos examinar el problema anterior para un nuevo valor de c, igual a $c' = c\sqrt{n'}/\sqrt{n}$ (entonces $\alpha_2 - \alpha_1$ puede escribirse en forma de $c'/\sqrt{n'}$). Por consiguiente, las potencias de $\pi_1(X', Y')$ y de $\pi_2(X', Y')$ en ese mismo punto $(\mathbf{P}_1, \mathbf{P}_2)$ serán aproximadamente iguales a

$$\beta_0\left(\sqrt{\frac{2}{\pi}}\, c'\right) = \beta_0\left(\sqrt{\frac{2n'}{\pi n}}\, c\right), \quad \beta_0\left(\sqrt{\frac{3}{\pi}}\, c'\right) = \beta_0\left(\sqrt{\frac{3n'}{\pi n}}\, c\right).$$

Igualando $\frac{2n'}{\pi n} = 1$, $\frac{3n'}{\pi n} = 1$, obtenemos los valores de $n' = \frac{\pi}{2}n$, $n' = \frac{\pi}{3}n$ (estos valores no dependen de c) para el número de observaciones que necesitamos realizar a fin de obtener con ayuda de los criterios π_1 y π_2, respectivamente, la misma potencia que para el criterio π_0 con n observaciones. Por ejemplo, para $n = 100$ observaciones con criterio π_0 necesitaremos, para obtener esos mismos resultados, $n' \approx 105$ observaciones con criterio π_2 y $n' \approx 157$ observaciones con criterio π_1.

Obtendríamos absolutamente otros resultados si hubiéramos verificado la homogeneidad para la familia $\mathscr{P} = \{\Phi_{0,\sigma^2}\}$. En este caso los criterios de signos y de Wilkoxon resultarían inconciliables. Más aún, el criterio de signos de nivel $1 - \varepsilon$ sería, en realidad, equivalente al criterio $\pi \equiv \varepsilon$ que

no depende de las muestras, ya que $\mathbf{M}(x_1 - y_1) = 0$ y $\mathbf{P}_1 \times \mathbf{P}_2(x_1 - y_1 > 0) = 1/2$ para cualquier par de distribuciones $\mathbf{P}_1$ y $\mathbf{P}_2$ de $\mathscr{P}$. Para este problema se podrían examinar otros criterios no paramétricos que utilizan las estadísticas r_i, por ejemplo, el criterio $\sum_{i=0}^{n_1} (r_{i+1} - r_i)^2$, $r_0 = 0$, $r_{n_1+1} = n_2$ que se asemeja por sus propiedades al criterio de Morán (§ 3.12).

5. Criterio χ^2 como criterio asintóticamente óptimo para verificar la homogeneidad según los datos agrupados. En este apartado supondremos que los datos en ambas muestras X e Y de volúmenes n_1 y n_2, respectivamente, están agrupados (véase el § 3.16). En este caso en vez de las muestras X e Y es posible utilizar los vectores $\nu = (\nu_1, \ldots, \nu_r)$ y $\mu = (\mu_1, \ldots, \mu_r)$ de las frecuencias de observaciones de las muestras X e Y, respectivamente, que cayeron en los intervalos $\Delta_1, \ldots, \Delta_r$ que definen la agrupación. Designemos por $\theta_i = (\theta_{i1}, \ldots, \theta_{ir})$, $i = 1, 2$, los vectores de las probabilidades de que las observaciones de la primera y la segunda muestras caigan en los intervalos $\Delta_1, \ldots, \Delta_r$, de modo que $\theta_{1l} = \mathbf{P}(x_j \in \Delta_l)$, $\theta_{2l} = \mathbf{P}(y_j \in \Delta_l)$. Las muestras aproximadas X e Y entonces pueden considerarse como muestras de las familias paramétricas $\mathbf{B}_{\theta_1}$ y $\mathbf{B}_{\theta_2}$, respectivamente. Ahora bien, el problema llega a ser paramétrico y podemos utilizar los resultados citados en el ejemplo 1 del párrafo precedente. De este ejemplo se deduce que si verificamos la hipótesis de homogeneidad $H_1 = \{\theta_1 = \theta_2\}$ en el caso en que el parámetro θ está localizado, o sea, los valores de θ_1 y θ_2 se sitúan en el entorno del punto $\theta_0 = (\theta_{01}, \ldots, \theta_{0r})$, entonces el criterio asintóticamente minimax de nivel asintótico $1 - \varepsilon$ para verificar H_1 frente a

$$H_2^b = \left\{ \sum_{l=1}^{r} \frac{(\theta_{1l} - \theta_{2l})^2}{\theta_{0l}} \geqslant \frac{b^2}{n_2} \right\}$$

tiene la forma

$$\sum_{l=1}^{r} \left(\frac{\nu_l}{n_1} - \frac{\mu_l}{n_2} \right)^2 \frac{n_1 n_2}{\nu_l + \mu_l} \geqslant h_\varepsilon,$$

donde h_ε es una cuantila del orden de $1 - \varepsilon$ de la distribución χ^2 con $r - 1$ grados de libertad. Este es precisamente el criterio χ^2 para verificar la homogeneidad según los datos agrupados.

En calidad de criterio asintóticamente equivalente puede ser considerado el criterio

$$\sum_{l=1}^{r} \nu_l \ln \frac{\nu_l}{n_1} + \sum_{l=1}^{r} \mu_l \ln \frac{\mu_l}{n_2} - \sum_{l=1}^{r} (\nu_l + \mu_l) \ln \frac{\nu_l + \mu_l}{n_1 + n_2} > \frac{h_\varepsilon}{2}.$$

§ 3. Problemas de regresión

1. Planteamiento del problema. En las aplicaciones a menudo surgen problemas referentes a las observaciones cuya distribución varía en distintos experimentos al cambiar algunos parámetros que caracterizan estos últimos. El conjunto de valores de los parámetros mencionados en el i-ésimo experimento, $i = 1, \ldots, n$ lo designaremos por

$$\mathbf{x}_i = (\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,r})$$

(así que r es la dimensión de los vectores $\mathbf{x}_i$). Los valores de $\mathbf{x}_{i,k}$ son determinados por el experimentador o por la naturaleza del fenómeno que se estudia. Designemos el vector $(\mathbf{x}_{1,k}, \ldots, \mathbf{x}_{n,k})$ por la letra X_k, y la matriz $\left(\frac{X_1}{X_r}\right) = (x_1^T, \ldots, x_n^T)$, por la letra X. Ahora bien, aquí, a distinción de lo expuesto anteriormente, X es una matriz del orden de $r \times n$ y puede ser un conjunto no aleatorio arbitrario de números cuya naturaleza no nos interesará. El vector de observaciones se designa por $Y = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$.

Los problemas de regresión están relacionados con la suposición de que las observaciones $\mathbf{y}_i$, en función del conjunto de parámetros $\mathbf{x}_i = (\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,r})$, tienen la forma

$$\mathbf{y}_i = \alpha_1 \mathbf{x}_{i,1} + \ldots + \alpha_r \mathbf{x}_{i,r} + \xi_i, \quad i = 1, \ldots, n, \tag{1}$$

donde $\alpha = (\alpha_1, \ldots, \alpha_r)$ son constantes desconocidas para nosotros, y $\xi_i \in \Phi_{0,\sigma^2}$ son constantes independientes.

La constante α_1 desempeña a menudo un papel especial, ya que en una serie de casos ésta separa en la representación (1) el sumando constante, lo cual corresponde a que en la matriz X se supone de antemano $X_1 = (1, \ldots, 1)$ $(\mathbf{x}_{i,1} \equiv 1)$. No haremos uso de esta suposición. Las variables aleatorias ξ_i se deben a los ruidos y fluctuaciones o a los errores de medición.

En forma matricial las relaciones (1) pueden escribirse del modo siguiente

$$Y = \alpha X + \xi. \tag{2}$$

La regresión que tiene la forma (1) y (2) se llama lineal (tanto respecto a α como respecto a X). En calidad de problemas de regresión pueden considerarse tanto el problema de estimación de los parámetros desconocidos α y σ^2, si se sabe que es válida (1), (2), como el problema de verificación de la propia hipótesis de que la representación (1), (2) tiene lugar. En ambos casos, como datos iniciales sirve la «muestra» (X, Y). El término «muestra» se utiliza aquí en un sentido más amplio que antes, designando con él el conjunto de resultados de observaciones que no tienen obligatoriamente la misma naturaleza. Además, recordemos que la primera de las dos «mues-

tras» X e Y puede ser no aleatoria. La matriz X se llama, a veces, *regresor* y el vector Y, *respuesta*.

El modelo de regresión (1), (2) es muy general si se tiene en cuenta que y_i depende del conjunto de parámetros. Suponiendo, por ejemplo, $x_{i,k} = \psi_k(z_i)$, donde $\psi_1, \ldots, \psi_r$ es un conjunto dado de funciones, y z_i son los valores del parámetro unidimensional, obtenemos el modelo

$$y_i = \alpha_1\psi_1(z_i) + \ldots + \alpha_r\psi_r(z_i) + \xi_i, \quad i = 1, \ldots, n, \tag{3}$$

de la regresión respecto a las funciones arbitrarias $\psi_1, \ldots, \psi_r$ (y, como antes, lineal respecto a α). Si $\psi_1(z) \equiv 1$, $\psi_2(z) \equiv z$ y $r = 2$, obtenemos el modelo de una *regresión lineal* elemental (unidimensional) (fig. 6).

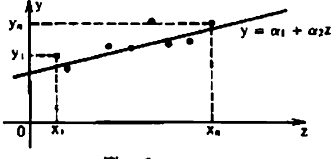


Fig. 6.

A distinción del modelo elemental, el modelo general (1), (2) se denomina, a veces, regresión *múltiple*. En general, como vemos, los problemas de regresión están relacionados con el estudio (existencia) de la dependencia funcional $y = \varphi(x)$ para una clase dada de funciones φ en los casos en que las observaciones de la variable y, para x dada, van acompañadas de «ruidos» en forma de desviaciones aleatorias.

Las filas $X_1, \ldots, X_r$ de la matriz X en (2) suelen elegirse de modo que sean linealmente independientes (de otro modo no podremos estimar las coordenadas de α). También seguiremos este convenio que significa que el rango de la matriz X es igual a r.

A veces es más cómodo tratar con los vectores ortogonales $X_1, \ldots, X_r$, o sea, con los vectores que satisfacen la condición $(X_i, X_j) = 0$, $i \neq j$, donde (a, b) significa el producto escalar. Si el conjunto inicial de vectores linealmente independientes (X_k) no posee tal propiedad, el mismo puede ser ortogonalizado introduciendo nuevos vectores:

$$\begin{aligned} X_1' &= X_1, \\ X_2' &= X_2 + a_{2,1}X_1, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_r' &= X_r + a_{r,r-1}X_{r-1} + \ldots + a_{r,1}X_1. \end{aligned} \tag{4}$$

Los coeficientes $a_{k,j}$ se deducen fácilmente de las condiciones de ortogonalidad $X_k' \perp X_j'$, $k \neq j$, así que, por ejemplo, $a_{2,1} = -\dfrac{(X_1, X_2)}{(X_1, X_1)}$. Las relacio-

nes (4) pueden ser escritas en forma de $X' = AX$, donde A es una matriz invertible triangular (con unidades que pasan por la diagonal principal). De aquí obtenemos $X = A^{-1}X'$, $Y = \alpha A^{-1}X' + \xi$. Hemos llegado al problema de regresión con coeficientes $\beta = \alpha A^{-1}$. El vector α se reconstruye de un modo evidente por β con ayuda de la igualdad $\alpha = \beta A$.

Para una regresión lineal elemental, la suposición acerca de la ortogonalidad de $X_1 = (1, \ldots, 1)$ y $X_2 = (z_1, \ldots, z_n)$ significa la suposición de $\sum z_i = 0$ que, evidentemente, puede ser satisfecha variando el comienzo de la lectura de la variable z.

2. Estimación de los parámetros. En lo sucesivo supondremos por doquier, que $r < n$ y que los vectores X_k, $k = 1, \ldots, r$, son linealmente independientes. La función de verosimilitud de la observación Y (con X dada) para la regresión (1), (2) es igual a

$$f_{\alpha,\sigma^2}(Y) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)_{i=1}^{n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \sum_{k=1}^{r}\alpha_k x_{i,k}\right)\right\}^2 =$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{|Y - \alpha X^2|}{2\sigma^2}\right\}. \qquad (5)$$

La función (5) depende del parámetro $\theta = (\alpha, \sigma^2)$. Nótese que si (5) se considera como función de verosimilitud no de una sola observación Y (o, (X, Y), sino de n observaciones $y_1, \ldots, y_n$, ella no corresponderá a la muestra de una familia paramétrica cualquiera. Las observaciones y_i se refieren a distintas distribuciones $\Phi_{\gamma_i, \sigma^2}$, $\gamma_i = \sum_{k=1}^{r} \alpha_k x_{i,k}$ que dependen de x_i. Por eso las consideraciones expuestas en los capítulos anteriores, donde se utilizó la misma distribución de los elementos de la muestra, aquí no se aplican directamente.

Así pues, examinaremos (5) como función de verosimilitud de la observación (X, Y). Hagamos uso del método de verosimilitud máxima. Directamente de (5) se deduce que la estimación de verosimilitud máxima $\alpha^* = \hat{\alpha}^*$ que maximiza $f_\theta(Y)$ respecto a α es la estimación que minimiza $|Y - \alpha X|^2$. Por eso en nuestro caso el método de verosimilitud máxima coincide con el «método de cuadrados mínimos».

Designemos por $\mathscr{A}[X]$ el subespacio tendido en los vectores $X_1, \ldots, X_r$. El mismo constituye una población de puntos en forma de αX cuando α recorre los valores de R^r. La dimensión de este espacio es r y en él sólo hay un punto $\beta = \alpha^* X$ que es el menos alejado de Y (fig. 7). El valor de β está unívocamente determinado por la condición de ortogonalidad $Y - \beta$ y $\mathscr{A}[X]$, o bien, que es lo mismo, por las r condiciones

$$(Y - \alpha^* X, X_k) = (Y - \alpha^* X)X_k^T = 0, \quad k = 1, \ldots, r.$$

En forma matricial estas condiciones pueden escribirse del modo siguiente: $(Y - \alpha^* X)X^T = 0$. De aquí hallamos

$$\alpha^* = YX^T(XX^T)^{-1}. \tag{6}$$

Aquí, la matriz inversa $(XX^T)^{-1}$ (del orden de $r \times r$) existe, ya que la matriz $D = XX^T$ está definida positivamente. En efecto, hemos visto que existe

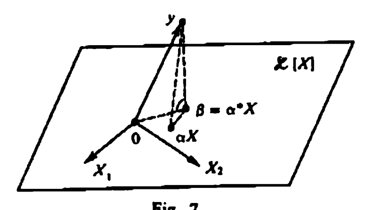


Fig. 7.

una matriz no degenerada A tal que las filas de la matriz $X' = AX$ son ortogonales. Por consiguiente, la matriz D puede ser escrita del modo siguiente:

$$XX^T = A^{-1}X'(X')^T(A^{-1})^T = A^{-1}B(A^{-1})^T,$$

donde $B = X'(X')^T$ es una matriz diagonal con los elementos

$$(X_i', X_j') = \begin{cases} |X_i'|^2 > 0 & \text{para } i = j, \\ 0 & \text{para } i \neq j. \end{cases}$$

Por lo tanto, B está definida positivamente, $aBa^T > 0$ para cualquier $a \in R^r$, $a \neq 0$. Poniendo $b = aA$, obtenemos $bDb^T = aAXX^TA^Ta^T = aBa^T > 0$ para cualquier $b \in R^r$, $b \neq 0$, que es lo que se necesitaba demostrar.

Si X_k son ortogonales, de (6) hallamos $\alpha_k^* = \dfrac{(Y, X_k)}{(X_k, X_k)}$.

El resultado (6) también puede ser obtenido derivando (5) respecto a α_k e igualando a cero las derivadas.

La diferencia $Y - \alpha^* X$ a veces se llama *resto*. Esta diferencia es ortogonal a $\mathscr{L}[X]$ y, al mismo tiempo, a cualquier vector $\gamma X \in \mathscr{L}[X]$, $\gamma \in R^r$. Si se adopta $\gamma = \alpha^* - \alpha$, de la igualdad $Y - \alpha X = Y - \alpha^* X + (\alpha^* - \alpha)X$ se deducirá

$$|Y - \alpha X|^2 = |Y - \alpha X|^2 + |(\alpha^* - \alpha)X|^2. \tag{7}$$

Hallemos ahora la e.v.m. para σ^2. De (5) se deduce que ésta será la misma estimación que para una familia normal (se puede volver a derivar (5) respecto a σ, igualando a cero la derivada), así que

$$(\widehat{\sigma^2})^* = \frac{1}{n}|Y - \alpha^* X|^2. \tag{8}$$

Pongamos

$$(\sigma^2)^* = \frac{1}{n-r}\,|Y - \alpha^* X|^2 = \frac{1}{n-r}\,(\widehat{\sigma^2})^*. \tag{9}$$

En lo sucesivo E_l significará una matriz unidad de orden l, $\sigma^* = \sqrt{(\sigma^2)^*}$.

Teorema 1. (6) y (9) *son las estimaciones eficientes no desplazadas e independientes de los parámetros* α y σ^2. *Además,*

$$(\alpha^* - \alpha)D^{1/2} \in \Phi_{0,\sigma^2 E_r},\ D = XX^T, \tag{10}$$

$$(n-r)(\sigma^2)^*/\sigma^2 = |Y - \alpha^* X|^2/\sigma^2 \in H_{n-r}. \tag{11}$$

Si X_k *son ortogonales,* α_k^* *son independientes,*

$$(\alpha_k^* - \alpha_k)|X_k| \in \Phi_{0,\sigma^2}. \tag{12}$$

Corolario 1. *De* (10) y (11) *se deduce que*

$$\frac{(\alpha^* - \alpha)D(\alpha^* - \alpha)^T}{(n-r)(\sigma^2)^*} = \frac{|(\alpha^* - \alpha)X|^2}{|Y - \alpha^* X|^2} \in F_{r,n-r}. \tag{13}$$

Sean $\bar{\alpha}$, $\bar{\alpha}^*$ *"subvectores" de dimensión* $l \leqslant r$ *de los vectores* α *y* α^* *formados por coordenadas de números fijos* $k_1, \ldots, k_l$, *y sea* $\overline{X}$ *una matriz formada por las filas* $X_{k_1}, \ldots, X_{k_l}$. *Entonces, si* X_k, $k = 1, \ldots, r$, *son ortogonales, entonces*

$$(\bar{\alpha}^* - \bar{\alpha})(\overline{X}\overline{X}^T)^{1/2} \in \Phi_{0,\sigma^2 E_l},\ (\alpha_k^* - \alpha_k)|X_k|/\sigma^* \in T_{n-r}. \tag{14}$$

Demostración del teorema 1. En vista de que $YX^T = \alpha XX^T + \xi X^T$, entonces

$$\alpha = (YX^T - \xi X^T)D^{-1},\quad \alpha^* - \alpha = \xi X^T D^{-1}. \tag{15}$$

La matriz de segundos momentos del vector $(\alpha^* - \alpha)D^{1/2}$ es igual a

$$\mathbf{M}D^{1/2}(\alpha^* - \alpha)^T(\alpha^* - \alpha)D^{1/2} =$$
$$= D^{1/2}D^{-1}X\mathbf{M}\xi^T\xi X^T D^{-1}D^{1/2} = \sigma^2 E_r.$$

Como las componentes de este vector son normales, ellas son independientes y $\frac{1}{\sigma^2}\,|(\alpha^* - \alpha)D^{1/2}|^2 \in \mathbf{H}_r$. Luego, en virtud de (7) y (9),

$$(n-r)(\sigma^2)^* = |Y - \alpha^* X|^2 = |\xi|^2 - |(\alpha^* - \alpha)X|^2.$$

Cerciorémonos ahora de que los vectores α^* e $Y - \alpha^* X$ (y, por consiguiente, α^* y σ^*) son independientes. En virtud de su normalidad es suficiente comprobar que los coeficientes de correlación entre sus componentes son iguales a cero o bien, que es lo mismo, que la matriz de segundos momentos centrales $\mathbf{M}(\alpha^* - \alpha)^T(Y - \alpha^* X)$ es igual a cero. Nótese que en virtud de (6),

$$\alpha^* X = YX^T(XX^T)^{-1}X = YX^T D^{-1}X,$$

y el vector $\alpha^* X$ se obtiene de Y mediante la proyección de Y sobre $\mathscr{A}[X]$. El operador de proyección, definido por la matriz $\Pi = X^T D^{-1} X$, posee propiedades evidentes: $\Pi^2 = \Pi$, $BX\Pi = BX$ para cualquier matriz B que tiene r filas. Por eso, en virtud de (15),

$$\mathbf{M}(\alpha^* - \alpha)^T (Y - \alpha^* X) = \mathbf{M} D^{-1} X \xi^T (\xi - \xi X^T D^{-1} X) = \\ = D^{-1} X \sigma^2 (E_n - \Pi) = 0.$$

Demostremos ahora (11). En virtud de (7),

$$|Y - \alpha^* X|^2 = |\xi|^2 - |(\alpha^* - \alpha) X|^2 = |\xi|^2 - |(\alpha^* - \alpha) D^{1/2}|^2,$$

donde $\frac{1}{\sigma^2} |\xi|^2 \in \mathbf{H}_n$, $\frac{1}{\sigma^2} |(\alpha^* - \alpha) D^{1/2}|^2 \in \mathbf{H}_r$ (véase (10)). La afirmación (11) será el corolario de estas relaciones y del lema 1.

Lema 1. *Si* $\eta = \eta_1 + \eta_2$, *donde* η_1 *y* η_2 *son independientes,* $\eta \in \mathbf{H}_n$, $\eta_1 \in \mathbf{H}_r$, *entonces* $\eta_2 \in \mathbf{H}_{n-r}$.

Demostración. Si se designa por $\varphi(t)$ la función característica de la distribución $\mathbf{H}_1$: $\varphi(t) = (1 + 2it)^{-1/2}$, entonces

$$\mathbf{M} e^{it\eta} = \varphi(t)^n = \varphi(t)^r \mathbf{M} e^{it\eta_2}.$$

Como $\varphi(t) \neq 0$ en el eje real, entonces $\mathbf{M} e^{it\eta_2} = \varphi(t)^{n-r}$. El lema queda demostrado.

El no desplazamiento de las estimaciones α^* y $(\sigma^2)^*$ se deduce con evidencia de (10), (11) ($\mathbf{M}\eta = l$ si $\eta \in \mathbf{H}_l$).

Nos queda demostrar la eficacia de la estimación $\theta^* = (\alpha^*, (\sigma^2)^*)$. Para esto debemos notar que la familia (5) pertenece al tipo exponencial, ya que (5) es representable en la forma (véase (2.15.1))

$$f_\theta(Y) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}(|Y|^2 - 2(Y, \alpha X) + |\alpha X|^2)\right\} = \\ = h(Y) \exp\left\{\sum_{k=1}^{r+1} a_k(\theta) U_k(Y) + V(\theta)\right\},$$

donde

$$h(Y) = (2\pi)^{-n/2}, \quad V(\theta) = -n \ln \sigma - \frac{1}{2\sigma^2} |\alpha X|^2,$$

$$a_k(\theta) = \frac{\alpha_k}{\sigma^2}, \quad U_k(Y) = (Y, X_k), \quad k = 1, \ldots, r,$$

$$a_{r+1}(\theta) = -\frac{1}{2\sigma^2}, \quad U_{r+1}(Y) = |Y|^2.$$

Como las condiciones de los teoremas 2.15.1 y 2.15.2 aquí se cumplen, la estadística $U = (U_1(X), \ldots, U_{r+1}(X))$ (y junto con ella también θ^*) es una estadística mínima suficiente completa. De aquí se desprende (véase el corolario 2.15.1) la eficacia de θ^*.

La afirmación (12) resulta con evidencia de (10), ya que para X_k ortogonales, la matriz $D^{1/2}$ es diagonal a los elementos $|X_k|$ dispuestos diagonalmente. El teorema queda demostrado.

Observación 1. Hotelling (véase [83]) demostró que $\mathbf{D}\alpha_k \geqslant \sigma^2/|X_k|^2$ y la igualdad se alcanza tan sólo en el caso cuando X_k son ortogonales. Ahora bien, al planificar un experimento para valores dados de $|X_k|$, la elección óptima del regresor X consiste en hacer ortogonales X_k.

Observación 2. Es interesante comparar la matriz de segundos momentos de la estimación θ^*, con la frontera inferior para las estimaciones no desplazadas, la cual se define, en virtud de la desigualdad multidimensional de Rao—Cramer, por la matriz $I^{-1}(\theta)$, donde $I(\theta)$ es la matriz de información de Fisher:

$$I(\theta) = \|I_{ij}(\theta)\|, \quad I_{ij}(\theta) = \mathbf{M}\theta \, \frac{\partial L}{\partial \theta_i} \cdot \frac{\partial L}{\partial \theta_j}, \quad L = L(Y;\ \theta) = \ln f_\theta(Y).$$

Aquí hemos adoptado $\theta_k = \alpha_k$, $k = 1, \ldots, r$, $\theta_{r+1} = \sigma^2$. Supongamos, para abreviar, que X_k son ortogonales. De la independencia de θ_k^* se deduce que la matriz $\mathbf{M}_\theta(\theta^* - \theta)^T(\theta^* - \theta)$ será diagonal a los elementos dispuestos diagonalmente:

$$\mathbf{M}_\theta(\alpha_k^* - \alpha)^2 = \frac{\sigma^2}{|X_k|^2}, \quad k = 1, \ldots, r,$$

$$\mathbf{M}_\theta((\sigma^2) - \sigma^2) = \mathbf{M}\left(\frac{\sigma^2 \chi^2_{n-r}}{n-r} - \sigma^2\right)^2 = \frac{\sigma^4}{n-r}\,\mathbf{M}(\chi_1^2 - 1)^2 = \frac{2\sigma^4}{n-r},$$

donde $\chi_l^2 \in \mathbf{H}_l$.

Por otro lado, para la matriz $I(\theta)$, en virtud de que

$$\frac{\partial L}{\partial \alpha_k} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{r} \alpha_j x_{ij}\right) x_{ik} = \frac{1}{\sigma^2}\,(Y - \alpha X)X_k^T,$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{r} \alpha_j x_{ij}\right)^2 = \frac{1}{2\sigma^2}\left(\frac{|Y - \alpha X|^2}{\sigma^2} - n\right),$$

hallamos, cuando $k = 1, \ldots, r$:

$$I_{kk}(\theta) = \mathbf{M}_\theta \frac{1}{\sigma^4} X_k(Y - \alpha X)^T(Y - \alpha X)X_k^T =$$

$$= \frac{1}{\sigma^4}\,\mathbf{M}X_k\xi^T\xi X_k^T = \frac{1}{\sigma^4}\,\mathbf{M}(\xi, X_k)^2 = \frac{|X_k|^2\mathbf{M}|\xi|^2}{\sigma^4} = \frac{|X_k|^2}{\sigma^2},$$

$$I_{r+1,r+1}(\theta) = \frac{1}{4\sigma^4}\,\mathbf{M}\left[\sum_{l=1}^{n}\left(\frac{\xi_l^2}{\sigma^2} - 1\right)\right]^2 = \frac{n}{2\sigma^4}, \quad I_{ij}(\theta) = 0$$

cuando $i \neq j$. Así que

$$I^{-1}(\theta) = \begin{Bmatrix} \frac{\sigma^2}{|X_1|^2} & 0 & 0 \\ \cdot & & \cdot \\ & \cdot & \cdot \\ 0 & \frac{\sigma^2}{|X_r|^2} & 0 \\ 0 \dots 0 & & \frac{2\sigma^4}{n} \end{Bmatrix}$$

Por lo tanto, en la desigualdad de Rao—Cramer,

$$\mathbf{M}_\theta(\theta^* - \theta)^T(\theta^* - \theta) \geqslant I^{-1}(\theta), \tag{16}$$

para las primeras r componentes de θ^* se alcanza la igualdad. Para la componente $r + 1$, la igualdad no puede alcanzarse (aunque asintóticamente, para $n \to \infty$, ambos miembros de (16) se comportan con igualdad), ya que la condición necesaria y suficiente del teorema 2.16.1A aquí no se cumple.

Observación 3. La suposición acerca de la normalidad de ε_i se vuelve poco importante para las afirmaciones (10)—(12), si n es grande (en (11) es mejor realizar la normalización y afirmar la proximidad a la ley normal).

Observación 4. El propio término "regresión" se refiere a la distribución conjunta de dos variables aleatorias ξ y η y significa la curva

$$g(x) = \mathbf{M}(\eta/\xi = x)$$

que también se llama regresión de η en ξ. Por ejemplo, si $(\xi, \eta) \in \Phi_{\gamma,\sigma^2}$, $\gamma = (\gamma_1, \gamma_2)$, $\sigma^2 = |\sigma_{ij}|$, $i, j = 1, 2$, entonces, como hemos visto en los capítulos anteriores, $g(x) = \gamma_2 + \frac{\sigma_{12}}{\sigma_{22}}(x - \gamma_1)$. Esta es una regresión lineal elemental.

Observación 5. La suposición $\xi_i \in \Phi_{0,\sigma^2}$ acerca de la igual distribución de ξ_i cuando se conoce σ^2, puede ser debilitada. Podemos considerar que $\xi_i \in \Phi_{0,\sigma_i^2}$, si σ_i son distintas y conocidas. En este caso, designando por σ la matriz diagonal $\sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix}$ e introduciendo nuevas variables $\xi' = \xi\sigma^{-1}$, $X' = X\sigma^{-1}$, $Y' = Y\sigma^{-1}$ (así que $\xi_i' = \xi_i/\sigma_i$, $x_i' = x_i/\sigma_i$, $y_i' = y_i/\sigma_i$), llegaremos al problema de regresión

$$Y' = \alpha X' + \xi'$$

en el que conocemos el vector de observaciones Y' y el regresor X', $\xi' \in \Phi_{0,E_n}$. Es fácil comprobar (el lector puede hacerlo personalmente) que es válido el siguiente análogo del teorema 1

Teorema 2. *La estimación*

$$\alpha^* = Y\sigma^{-2}X^T(D')^{-1}, \quad D' = X\sigma^{-2}X^T,$$

es la estimación eficiente no desplazada de α,

$$(\alpha^* - \alpha)(D')^{1/2} \in \Phi_{0,E_r},$$

$$|Y' = \alpha^* X'|^2 = \sum_{i=1}^{n} \frac{\left(y_i - \sum_{k=1}^{r} \alpha_k^* x_{ik}\right)^2}{\sigma_i^2} \in \mathbf{H}_{n-r}.$$

Recurramos de nuevo al teorema 1. Las relaciones (10)—(12) establecidas en este teorema permiten construir conjuntos confidenciales tanto para distintas coordenadas de θ como para el vector θ en total. Por ejemplo,

$$\mathbf{P}_\theta\left(\frac{(n-r)(\sigma^2)^*}{h_{\varepsilon/2}^{(2)}} < \sigma^2 < \frac{(n-r)(\sigma^2)^*}{h_{\varepsilon/2}^{(1)}}\right) = 1 - \varepsilon, \tag{17}$$

y si X_k son ortogonales, entonces

$$\mathbf{P}_\theta\left(|\alpha_k - \alpha_k^*| < \frac{t_{\varepsilon/2}\sigma^*}{|X_k|}\right) = 1 - \varepsilon, \tag{18}$$

donde $\mathbf{T}_{n-r}((-t_{\varepsilon/2}, t_{\varepsilon/2})) = 1 - \varepsilon$, $\mathbf{H}_{n-r}((h_{\varepsilon/2}^{(1)}, h_{\varepsilon/2}^{(2)})) = 1 - \varepsilon$.

Supongamos que X_k son ortogonales. Designemos por $\bar{\alpha}$ el "subvector" del vector α, definido en el corolario 1. En virtud del teorema 1 es natural construir el conjunto confidencial para $\bar{\alpha}$ valiéndose de la relación

$$\frac{|(\bar{\alpha} - \bar{\alpha}^*)\overline{X}|^2}{(n-r)(\sigma^2)^*} < f_\varepsilon. \tag{19}$$

El valor de f_ε, correspondiente al nivel disponible $1 - \varepsilon$, se determina de manera conocida (véase el capítulo 3), o sea, mediante la distribución de Fisher $\mathbf{F}_{l,n-r}$ con l, $n - r$ grados de libertad.

Si se conoce σ^2, el intervalo confidencial será definido por la relación

$$|(\bar{\alpha} - \bar{\alpha}^*)\overline{X}|^2 < \sigma^2 h_\varepsilon, \tag{20}$$

donde h_ε corresponde a la distribución $\mathbf{H}_l$.

En los problemas de regresión puede resultar que también sea necesario estimar el valor de la superficie de regresión $y = \alpha z^T$ en un nuevo punto dado de antemano, $z = (z_1, \ldots, z_r) \in R^r$. Pongamos $y^* = \alpha^* z^T$. Entonces, como antes, hallamos

$$y^* - y = (\alpha^* - \alpha)z^T = \xi X^T D^{-1} z^T \in \Phi_{0,d^2},$$

$$d^2 = \sigma^2 z D^{-1} z^T, \quad \frac{y^* - y}{d\sigma^*} \in \mathbf{T}_{n-r}.$$

Esto da la posibilidad de construir los intervalos confidenciales para y.

Cabe señalar que la determinación de la región confidencial para la superficie de regresión es "en general" un problema más complejo (compárese con [30]). La población de las superficies que entran en el conjunto confidencial será determinada por el conjunto confidencial para θ construido, por ejemplo con ayuda de (10), (11) (véase el § 3.8). Esto se expone más detalladamente en [30].

3. Verificación de las hipótesis con respecto a la regresión lineal. Aquí toquemos dos tipos de problemas.

1) Supongamos que sabemos que la representación (1), (2) tiene lugar. Se necesita verificar la hipótesis de que θ es igual al valor dado de θ' o que el conjunto de coordenadas $\theta_{k_1}, \ldots, \theta_{k_l}$ es igual al conjunto de $\theta'_{k_1}, \ldots, \theta'_{k_l}$, mientras que las demás coordenadas se desconocen.

El criterio para verificar tales hipótesis ha de construirse con ayuda de los conjuntos confidenciales (17)—(20) (véase el § 3.8). Supongamos, por ejemplo, que se necesita verificar la hipótesis H_1 la independencia de Y respecto a X para una regresión lineal elemental, o sea, la hipótesis $H_1 = \{\alpha_2 = 0\}$. Entonces, de (18) (o de (14)) obtenemos el criterio de nivel $1 - \varepsilon$ que rechaza H_1 si

$$|\alpha_2^*| \geqslant t_{\varepsilon/2}\sigma^*/|X_2|. \tag{21}$$

En el caso general de la regresión (1) con X_k ortogonales, la hipótesis de independencia de Y respecto a X tendrá la forma $H_1 = \{\bar{\alpha} = 0\}$, donde $\bar{\alpha} = (\alpha_2, \ldots, \alpha_r)$, $x_{i1} \equiv 1$, y para su verificación se puede aprovechar el criterio

$$\frac{|\bar{\alpha}^* \overline{X}|^2}{(n - r)(\sigma^2)^*} \geqslant f_\varepsilon, \tag{22}$$

donde $\overline{X}$ y f_ε están definidas en (19) para $l = r - 1$.

También se pueden utilizar los enfoques del § 3.15, donde fue examinada la verificación de la pertenencia de la muestra a una subfamilia paramétrica. Entonces llegaremos al criterio de relación de verosimilitud, el cual, desde cierto punto de vista, será semejante a (22). Si se conoce σ^2, entonces, el c.r.v. para verificar $H_1 = \{\bar{\alpha} = 0\}$ tendrá la forma

$$\sigma^{-2}|\bar{\alpha}^* \overline{X}|^2 > h_\varepsilon,$$

donde h_ε es la cuantila $\mathbf{H}_{r-1}$ de orden $1 - \varepsilon$. Este criterio será minimax (véase los §§ 3.9 y 3.10) para las alternativas correspondientemente separadas.

2) Verificación de la hipótesis de que en la muestra (X, Y) está presente la propia regresión (1), (2). Por estas palabras entendemos la hipótesis de que para α y σ cualesquiera tiene lugar la representación (1), (2), o sea, para α y σ cualesquiera es válida $\sigma^{-1}(Y - \alpha X) \in \Phi_{0,E_n}$. Este es el proble-

ma de pertenencia de Y a una familia paramétrica. Pero como ya hemos señalado, las observaciones en Y no están igualmente distribuidas. Para reducir el problema al caso de distribuciones igualmente distribuidas (véase el § 3.17), haremos uso de la afirmación siguiente, que completa el teorema 1. Consideraremos que X_k son ortogonales.

Teorema 3. *Sea C cualquier matriz ortogonal de orden $n \times n$ que contiene, en calidad de primeras r columnas, las columnas de la matriz $X^T D^{-1/2}$. Entonces, el vector $\delta = (Y - \alpha^* X)C$ tiene coordenadas independientes que poseen la propiedad $\delta_1 = \ldots = \delta_r = 0$, $\delta_i \in \Phi_{0,\sigma^2}$, $i = r + 1, \ldots, n$.*

Ahora bien, el problema se reduce a la verificación de la hipótesis de pertenencia de la muestra $\delta_{r+1}, \ldots, \delta_n$, de volumen $n - r$, a la familia Φ_{0,σ^2} en términos generales (r observaciones se utilizaron para estimar α). Este problema fue examinado en el § 3.17. Para obtener los valores de δ_i es necesario, basándose en las muestras X e Y, calcular sucesivamente los valores de α^*, $Y - \alpha^* X$ y aplicar a $Y - \alpha^* X$ cualquier transformación C dotada de las propiedades indicadas en el teorema 3. Si se conoce σ, llegaremos al problema de verificación de la hipótesis simple de pertenencia de Φ_{0,σ^2}. No obstante, en este caso, para verificar la hipótesis que nos interesa también se puede utilizar el teorema 1, en virtud del cual

$$(n - r)(\sigma^2)^*/\sigma^2 \in \mathbf{H}_{n-r}.$$

Demostración del teorema 3. Si $Z \perp \mathscr{L}[X]$, entonces, las primeras r coordenadas del vector ZC forman el vector $ZX^T D^{-1/2} = 0$. Como $(Y - \alpha^* X) \perp \mathscr{L}[X]$ y $\delta = (Y - \alpha^* X)C$, de aquí resulta que $\delta_1 = \ldots = \delta_r = 0$. Seguidamente,

$$\delta = (Y - \alpha X)C - (\alpha^* - \alpha)XC = \eta - \bar{\eta} D^{-1/2} XC,$$

donde $\eta = \xi C$, $\bar{\eta} = (\eta_1, \ldots, \eta_r) = (\alpha^* - \alpha)D^{1/2} = \xi X^T D^{-1/2}$ y, por consiguiente, δ es el resultado de la transformación lineal sobre η,

$$|\delta|^2 = |Y - \alpha^* X|^2 = |\xi|^2 - |(\alpha - \alpha^*)X|^2 = \sum_{i=1}^{n} \eta_i^2 - |\bar{\eta}|^2 = \sum_{i=r+1}^{n} \eta_i^2,$$

así que $\sum_{i=r+1}^{n} \delta_i^2 = \sum_{i=r+1}^{n} \eta_i^2$. Esto sólo es posible en el caso cuando $(\delta_{r+1}, \ldots, \delta_n)$ es el resultado del giro del vector $(\eta_{r+1}, \ldots, \eta_n)$, o bien, que es lo mismo, el resultado de la transformación ortogonal sobre $(\eta_{r+1}, \ldots, \eta_n)$. En vista de que $\sigma^{-1} \in \Phi_{0,E_n}$, el teorema queda demostrado.

Ejemplo 1. En este ejemplo describiremos el aspecto matemático de un experimento físico con cuya ayuda fue descubierto el efecto de desintegración del mesón φ en dos mesones π (véase [85]). El resultado obtenido tiene carácter estadístico y en él se utilizó, en esencia, el modelo de regresión.

La investigación se refiere al estudio de la interacción de los electrones (e^-) y los positrones (e^+) en los haces que vienen al encuentro. Si la energía total de estas partículas $2E$ se encuentra en el entorno del punto $2E_0 = 1019{,}6$ MeV (fig.8), entonces, al producirse el "choque", de las mismas, como resultado de la acción mutua se forman (a la par con otras) partículas de dos tipos: mesones φ y mesones π. La probabilidad de surgimiento de pares de mesones π durante la interacción de e^+ y e^- conforme a la energía E, se describe con gran precisión por medio de la función lineal que presentamos en forma de (hipótesis H_1)

$$p_1^{\pi\pi}(E) = a_0 + a_1x_1 \quad x = E - E_0, \tag{23}$$

donde a_0, a_1 se desconocen.

Fue planteada la suposición (hipótesis H_2) de que al desintegrarse los mesones φ generados, también pueden aparecer pares de mesones π. Prácticamente es imposible revelar este efecto de un modo directo, ya que se ha establecido que tal fenómeno, si ocurre, se produce muy raramente: no más de una vez en 10^4 desintegraciones de mesones φ. No obstante, gracias al efecto de interferencia de este canal adicional de engendramiento de mesones π, con el canal principal, la probabilidad de que se produzcan dichas partículas será igual no a (23) sino a

$$p_2^{\pi\pi}(E) = [a_0 + a_1x]\left[1 + \frac{b_0 + b_1x + b_2x^2}{x^2 + d^2}\right] \tag{24}$$

(al igual que en (23), ésta es una aproximación muy exacta de una fórmula más compleja, basada en el hecho de que el intervalo de variación que se examina, o sea, $x = E - E_0$, es pequeño en comparación con E_0). En esta igualdad, los coeficientes b_i, al igual que a_i, se desconocen, pero d se conoce.

Para establecer cuál de las dos relaciones, (23) ó (24), tiene lugar en realidad, se ejecutaron $n = 20$ experimentos con distintos valores de energía $E_1, \ldots, E_{20}$.

Los resultados de los experimentos (véase la tabla 1 y la fig. 8) son

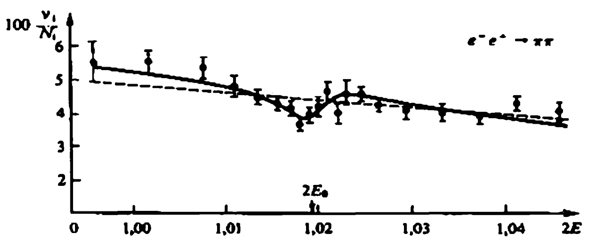


Fig. 8. Las curvas representan las estimaciones de las líneas de regresión para las hipótesis H_1 y H_2.

las cantidades N_i, $i = 1, \ldots, 20$ de interacciones de e^+ y e^-, y las cantidades ν_i de pares de mesones π engendrados con energía E_i. En cada uno de los experimentos efectuados, los números N_i y ν_i son bastante grandes (N_i es del orden de 10^4). En vista de que cuando N_i, es fijo, el número ν_i de pares de mesones π tiene distribución de Bernoulli $\mathbf{B}_{p_i}^{N_i}$ ($p_i = p_1^{\pi\pi}(E_i)$ en la hipótesis H_1, y $p_i = p_2^{\pi\pi}(E_i)$ en la hipótesis H_2), entonces, utilizando la aproximación normal, podemos considerar, con derecho, que tiene lugar la representación

$$y_i = \frac{\nu_i}{N_i} = p_i + \xi_i, \quad \xi_i \in \Phi_{0,\sigma_i^2}$$

Tabla 1. Tabla de los datos experimentales

Número del experimento i	E_i, MeV	N_i	ν_i	Número del experimento i	E_i, MeV	N_i	ν_i
1	497,75	6 960	384	11	510,40	14 322	716
2	500,65	7 908	435	12	510,92	13 470	568
3	503,65	8 102	432	13	511,39	12 008	569
4	505,40	22 259	1080	14	512,17	23 951	1117
5	506,62	16 938	765	15	513,20	27 796	1185
6	507,66	21 728	951	16	514,62	37 771	1539
7	508,40	14 014	603	17	516,58	25 902	1036
8	508,90	13 793	545	18	518,64	27 857	1057
9	509,40	14 075	615	19	520,61	23 228	989
10	509,90	14 867	691	20	522,88	26 482	1066

(en el sumando ξ_i también entran los ruidos eventuales (fondo)). En virtud de (23) y (24) tendremos dos posibles variantes de regresión:

$$p_i = \sum_{k=0}^{1} \alpha_k \psi_k(x_i), \quad \psi_k(x) = x^k, \ k = 0, 1 \tag{25}$$

(hipótesis H_1) y

$$p_i = \sum_{k=0}^{3} \alpha_k \psi_k(x_i), \quad \psi_k(x) = \frac{x^k}{x^2 + d^2}, \quad k = 0, 1, 2, 3 \tag{26}$$

(hipótesis H_2).

Al variar las hipótesis, los valores de σ_i^2 cambian muy poco; éstos pueden ser apreciados muy exactamente y podemos considerar que son conocidos. Entonces, basándose en el teorema 2, la distribución de la estadística

$$\chi^2 = \left| Y' - \sum_k \overset{*}{\alpha}_k \varphi_k \right|^2 = \sum_{l=1}^{n} \left(y_l - \sum_k \overset{*}{\alpha}_k \varphi_k(x_l) \right)^2 \Big/ \sigma_l^2 \tag{27}$$

será $\mathbf{H}_{n-r}$, donde r es el número de parámetros sujetos a estimación α_k ($r = 2$ en la hipótesis H_1, y $r = 4$ en la hipótesis H_2).

Tras realizar los cálculos necesarios conforme a las recomendaciones del teorema 2, obtendremos, para la estadística (27), los valores siguientes: en el primer caso ($r = 2$) $\chi_1^2 = 36{,}8$, y en el segundo ($r = 4$) $\chi_2^2 = 19{,}0$. Los niveles significativos realmente alcanzables (véase el § 3.4) del criterio $\chi^2 > c$ para verificar las hipótesis H_1 y H_2 (como principales) constituirán $H_{18}((0,\ 36{,}8)) = 0{,}9944$ y $H_{16}((0,\ 19{,}0)) = 0{,}731$.

Con otras palabras, la suposición de que falta el canal adicional de engendramiento de pares de mesones π es rechazada por el criterio fundado en la estadística χ^2 con nivel de significación igual, por ejemplo, a 0,99. Al mismo tiempo, la suposición acerca de la existencia de este canal concuerda bien con los resultados experimentales.

Hablando más exactamente, en este problema deberíamos verificar dos hipótesis paramétricas compuestas, correspondientes a las suposiciones (25) y (26) para los valores de las probabilidades de aparición de pares de mesones π. Si utilizamos el criterio de relación de verosimilitud, éste, como es fácil comprobar, se basará en la diferencia de las estadísticas χ^2 correspondientes a los modelos (25) y (26) y, por lo tanto, sus resultados serán aproximadamente los mismos.

4. Estimación y verificación de las hipótesis al existir relaciones lineales. Examinemos, como antes, la regresión lineal (1), (2), pero suponiendo que las coordenadas del vector α están ligadas mediante $s < r$ relaciones lineales

$$\sum_{k=1}^{r} \alpha_k a_{kl} = c_l, \quad l = 1, \ldots, s.$$

En forma matricial estas relaciones pueden escribirse del modo siguiente:

$$\alpha A = c, \tag{28}$$

donde A es una matriz de orden $r \times s$. Supongamos que A es de rango s. En este caso podríamos expresarlas s variables (digamos, $\alpha_{r-s+1}, \ldots, \alpha_r$) a través de las demás (o sea, a través de $\alpha_1, \ldots, \alpha_{r-s}$), sustituir los valores obtenidos en (1), (2) y volver a obtener el problema estándar de regresión lineal (pero con regresor modificado).

Pero para la exposición ulterior trataremos de resolver este problema de un modo algo distinto. Recurramos a la demostración del teorema 1. El subespacio $\mathscr{A}$ de valores α, definido por las relaciones (28), separa en $\mathscr{L}[X]$ el subespacio de dimensión s y de valores αX, el cual designaremos por $\mathscr{L}_A[X]$. Es evidente que la estimación $\alpha \in \mathscr{A}$ ahora puede efectuarse a base de los mismos procedimientos que hemos utilizado en el teorema 1. La estimación necesaria $\alpha_A^* \in \mathscr{A}$ será determinada, al igual que en el teorema 1, con ayuda de la proyección $\alpha_A^* X$ del vector Y sobre $\mathscr{L}_A[X]$. Ahora

bien, a la par con la relación $(Y - \alpha^* X) \perp \mathscr{L}[X]$ tendremos la relación $(Y - \alpha_A^* X) \perp \mathscr{L}_A[X]$ que define unívocamente α_A^*. Para obtener el propio valor de α_A^* es más cómodo hacer uso del enfoque analítico, o sea, aplicar el método de multiplicadores indeterminados de Lagrange para encontrar $\min_\alpha |Y - \alpha X|^2$ a condición de que $\alpha A = c$. Para esto debemos resolver las ecuaciones

$$\alpha A = c, \quad \frac{\partial}{\partial \alpha}\left[|Y - \alpha X|^2 + \lambda(\alpha A - c)^T\right] = 0 \tag{29}$$

(utilizamos los multiplicadores $\lambda_1, \ldots, \lambda_s$ que forman el vector λ y que corresponden a las condiciones (28)). En vista de que $|Y - \alpha X|^2 = (Y - \alpha X)(Y - \alpha X)^T$, la segunda de las ecuaciones (29) adoptará la forma siguiente:

$$-2YX^T - 2\alpha XX^T + \lambda A^T = 0.$$

De aquí hallamos

$$\alpha_A^* = YX^T D^{-1} - \frac{1}{2}\lambda A^T D^{-1} = \alpha^* - \frac{1}{2}\lambda A^T D^{-1}.$$

En virtud de (29), $c = \alpha_A^* A = \alpha^* A - \frac{1}{2}\lambda A^T D^{-1} A$. Como la matriz D está definida positivamente, y el rango de A es s, el rango de la matriz $B = D^{-1/2}A$ también será s, y la matriz $B^T B = A^T D^{-1} A$ también estará positivamente definida (véase el punto 1). Por consiguiente,

$$-\frac{1}{2}\lambda = (c - \alpha^* A)D_A,$$
$$\alpha_A^* = \alpha^* + (c - \alpha^* A)D_A A^T D^{-1}, \tag{30}$$

donde suponíamos, para abreviar, $D_A = [A^T D^{-1} A]^{-1}$.

El lector puede comprobar que hemos obtenido la e.v.m. del parámetro α a condición de que $\alpha A = c$. Ese mismo resultado (30) también se puede obtener de las consideraciones geométricas, utilizando las relaciones $\alpha_A^* X \in \mathscr{L}_A[X]$ y la ortogonalidad

$$\begin{gathered}(Y - \alpha_A^* X) \perp \mathscr{L}_A[X], \\ (\alpha_A^* - \alpha^*)X = [(Y - \alpha^* X) - (Y - \alpha_A^* X)] \perp \mathscr{L}_A[X].\end{gathered} \tag{31}$$

Recurramos ahora al problema de *verificación de las hipótesis lineales.* La hipótesis H_1 respecto al parámetro α se llamará hipótesis *lineal* si su forma es $H_1 = \{\alpha A = c\}$, donde las matrices A y c han sido definidas anteriormente.

Inmediatamente podemos señalar que introduciendo el nuevo parámetro $\beta = \alpha A_c$, donde A_c es cualquier matriz no degenerada, cuyas s primeras columnas coinciden con A, reduciremos el problema a la regresión

$$Y = \beta X' + \xi, \quad X' = A_c^{-1} X,$$

y a la verificación de la hipótesis $\bar{\beta} = c$, $\bar{\beta} = (\beta_1, \ldots, \beta_s)$ (véase el punto 2).

También es natural partir de las consideraciones siguientes. Cuanto más se distinga αA de c, tanto más lejos permanecerá αX de $\mathscr{L}_A[X]$ y tanto más se distinguirán los puntos αX y $\alpha^* X$ de $\alpha_A^* X \in \mathscr{L}_A[X]$. Por eso es natural suponer que la base del criterio para verificar H_1 es la distancia que separa $\alpha_A^* X$ de $\alpha^* X$. Si la hipótesis H_1 es cierta, entonces, en virtud de (31),

$$|(\alpha_A^* - \alpha^*)X|^2 = |Y - \alpha_A^* X|^2 - |Y - \alpha^* X|^2. \tag{33}$$

En virtud de (30) (sustituyendo c por αA), $\alpha_A^* - \alpha^*$ es el resultado de la transformación lineal sobre $\alpha - \alpha^*$. Por eso $(\alpha_A^* - \alpha^*)X$ no depende de $Y - \alpha^* X$ (véase el teorema 1).

Seguidamente, en virtud de (30),

$$\begin{aligned}|(\alpha_A^* - \alpha^*)X|^2 &= (\alpha_A^* - \alpha^*)XX^T(\alpha_A^* - \alpha^*)^T = \\ &= (c - \alpha^* A)D_A(c - \alpha^* A) = (\alpha^* - \alpha)AD_A A^T(\alpha^* - \alpha)^T.\end{aligned} \tag{34}$$

En vista de que

$$(\alpha^* - \alpha)A = \xi X^T D^{-1} A \in \Phi_{0,\sigma^2 A^T D^{-1} A} = \Phi_{0,\sigma^2 D_A^{-1}},$$

en virtud de (34) y del § 2.2 (punto 4)

$$\frac{1}{\sigma^2}|(\alpha_A^* - \alpha^*)X|^2 \in \mathbf{H}_s. \tag{35}$$

De lo dicho y del teorema 1 resulta que

$$\frac{|(\alpha_A^* - \alpha^*)X|^2}{|Y - \alpha^* X|^2} = \frac{|Y - \alpha_A^* X|^2}{|Y - \alpha^* X|^2} - 1 \in \mathbf{F}_{s,n-r}. \tag{36}$$

Las relaciones (35) y (36) nos permiten construir los criterios (basados en la utilización del alejamiento de $\alpha^* X$ respecto a $\alpha_A^* X$) para verificar la hipótesis H_1 en los casos cuando σ^2 se conoce y se desconoce, respectivamente (véase el capítulo 3).

Cabe señalar que H_1 es la hipótesis de pertenencia de α a una subfamilia paramétrica (al existir el parámetro obstaculizador σ^2, si σ^2 se desconoce), y las estadísticas (35) y (36) no son otra cosa sino las estadísticas de la relación de verosimilitud (véanse los §§ 3.10 y 3.15). En efecto, supongamos, por ejemplo, que desconocemos σ^2. Entonces (véanse (5) y (8)),

$$\sup_{\alpha,\sigma} f_\theta(Y) = \sup_{\alpha,\sigma}(\sqrt{2\pi}\,\sigma)^{-n}\exp\left\{-\frac{|Y - \alpha X|^2}{2\sigma^2}\right\} =$$

$$= (\sqrt{2\pi}\,\hat{\sigma}^*)^{-n}\exp\left\{-\frac{|Y - \alpha^* X|^2}{2(\sigma^2)^*}\right\} = \left(\sqrt{2\pi}\,\frac{|Y - \alpha^* X|}{n}\right)^{-n} e^{-n/2}.$$

El valor de $\sup\limits_{\alpha \in \mathscr{A},\sigma} f_\theta(Y)$ se calcula exactamente igual. Sólo es preciso señalar

que la e.v.m. para α, en el caso de $a\alpha \in \mathscr{A}$, será α_A^*, y la e.v.m. para σ^2 será igual, así como en (8), a $\frac{1}{2}$ $|Y - \alpha_A^* X|^2$. Por eso

$$\sup_{\sigma \in \mathscr{A}_\sigma} f_\theta(Y) = \left(\sqrt{2}\,\frac{|Y - \alpha_A^* X|}{n}\right)^{-n} e^{-n/2},$$

$$\frac{\sup\limits_{\alpha \in \mathscr{A}, \sigma} f_\theta(Y)}{\sup\limits_{\alpha, \sigma} f_\theta(Y)} = \frac{|Y - \alpha^* X|^n}{|Y - \alpha_A^* X|^n}$$

y, por consiguiente, la estadística del criterio de relación de verosimilitud equivale a (36).

Si σ^2 se conoce, como base del criterio para verificar H_1 se puede adoptar la relación (35). Análogamente a lo expuesto más arriba, el lector puede convencerse de que el resultado obtenido también es el criterio de la relación de verosimilitud. Como este criterio es invariable respecto a la sustitución del parámetro (véase el § 3.10), entonces, en virtud de la advertencia y las afirmaciones de los §§ 3.9 y 3.10, se puede afirmar que el c.r.v.

$$|(\alpha_A^* - \alpha^*)X|^2 > \sigma^2 h_\varepsilon,$$

donde h_ε es la cuantila de orden $1 - \varepsilon$ de la distribución $\mathbf{H}_s$, la cual constituirá el criterio minimax de nivel $1 - \varepsilon$ para verificar H_1 frente a las alternativas separadas respectivamente.

Lo dicho más arriba y los resultados de los capítulos 2 y 3 (en particular el § 3.15) dan razones para considerar que los criterios (36), al igual que la estimación (30), también poseen propiedades de optimización. Aquí no nos detendremos más detalladamente en este material. Una exposición más completa de los problemas de regresión se ofrece en [83].

§ 4. Análisis de varianza

Los problemas de análisis de varianza que se exponen en este párrafo pertenecen, en su esencia, a los problemas de regresión. En los últimos de ellos hemos estudiado la dependencia de las observaciones del factor numérico x que podía adoptar cualesquiera valores dados de antemano $x_1, \ldots, x_n$, y a cada uno de ellos le correspondía una sola observación. En los problemas de análisis de varianza suele estudiarse la influencia que ejercen únicamente los factores discretos (uno, dos o más) que pueden tomar exclusivamente un número finito de valores. Para cada uno de estos valores disponemos de un conjunto de observaciones (de una muestra). El análisis de varianza une un grupo de procedimientos estadísticos basados en el análisis de las desviaciones estándar y destinadas a verificar diversas hipótesis y estimar los parámetros relacionados con la influencia de los factores. Los fundamentos del análisis de varianza fueron establecidos por Fisher.

1. Problemas de análisis de varianza como problemas de regresión. El caso de un factor. Supongamos que se dan r muestras independientes

$$Y_1 = (y_{11}, \ldots, y_{1n_1}), \ldots, Y_r = (y_{r1}, \ldots, y_{rn_r})$$

de volúmenes $n_1, \ldots, n_r$ de las poblaciones normales: $Y_k \in \Phi_{\alpha_k,\sigma^2}$. Se supone que las observaciones Y_k, $k = 1, \ldots, r$ se han realizado con diferentes valores de cierto factor cuya importancia nos interesa y que la influencia de este factor se refleja en el valor de la media α_k. Se supone, además, que el valor de la varianza σ^2 es el mismo para todas las muestras y, por regla general, es desconocido. Los problemas de análisis de varianza comprenden la verificación de las hipótesis referentes a los valores $\alpha_1, \ldots, \alpha_r$ y, en particular, de la hipótesis acerca de la homogeneidad de $\alpha_1 = \ldots = \alpha_r = \alpha$ (en el § 1 hemos examinado este último problema), así como las estimaciones de los parámetros α_k y de su variabilidad.

Al igual que los problemas de regresión, el análisis de varianza se aplica ampliamente, sobre todo en la sociología, la agricultura, la biología y la medicina. En calidad de un problema muy típico para aplicar los métodos del análisis de varianza se puede nombrar, por ejemplo, el problema de aclaración de la dependencia que existe entre el contenido de colesterina en la sangre de una persona y su profesión.

Los problemas de análisis de varianza enunciados anteriormente son casos particulares de los problemas de regresión lineal. En efecto, las observaciones y_{ki} pueden representarse en la forma

$$y_{ki} = \alpha_k + \xi_{ki}, \ \xi_{ki} \in \Phi_{0,\sigma^2}, \ k = 1, \ldots, r, \ i = 1, \ldots, n_k. \quad (1)$$

Formemos el vector

$$Y = ((y_{11}, \ldots, y_{1n_1}; y_{21}, \ldots, y_{2n_2}; \ldots; y_{r1}, \ldots, y_{rn_r})$$

y el vector ξ observando esa misma regla. Entonces, las relaciones (1) pueden ser escritas en la forma matricial $Y = \alpha X + \xi$, donde X es una matriz de dimensión $r \times n$, $n = n_1 + \ldots + n_r$ que tiene la forma siguiente:

$$X = \left(\begin{array}{cccc|cccc|c|cccc}
1 & 1 & \ldots & 1 & 0 & 0 & \ldots & 0 & & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 & & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & & 0 & 0 & \ldots & 0 \\
\cdots & & & & \cdots & & & & & \cdots & & & \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & & 1 & 1 & \ldots & 1
\end{array}\right).$$

Es evidente que las filas de esta matriz (vectores X_j) son ortogonales. La hipótesis $H_1 = \{\alpha_1 = \alpha_2 = \ldots = \alpha_r\}$ puede escribirse del siguiente modo:

$$\alpha A = 0,$$

donde A es una matriz de dimensión $r \times (r-1)$:

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ -1 & -1 & \dots & -1 \end{pmatrix}.$$

Es evidente que el rango de A es $r-1$.

Vemos que la verificación de la hipótesis principal H_1 del análisis de varianza no es otra cosa sino el problema de verificación de la hipótesis lineal para la regresión.

Vamos a aclarar qué son las estimaciones eficientes para α y σ^2 halladas en el teorema 3.1. En nuestro caso $|X_k|^2 = n_k$, la matriz $D = XX^T$ de orden $r \times r$ tiene la forma

$$D = \begin{pmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & n_r \end{pmatrix},$$

$$\alpha_k^* = \frac{(Y, X_k)}{(X_k, X_k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} y_{ki} = \bar{y}_{k.}, \tag{3}$$

$$(n-r)(\sigma^2)^* = |Y - \alpha^* X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \bar{y}_{k.})^2 = Q_2(Y).$$

En este caso, $\alpha_1^*, \dots, \alpha_r^*, (\sigma^2)^*$ son independientes. Los intervalos confidenciales para los parámetros α, σ^2, así como sus funciones, se construyen al igual que en el § 3.

Para verificar la hipótesis lineal (2) también debemos calcular la e.v.m. α_A^* al existir la condición (2) (véase el punto 4 del párrafo anterior). Aquí, el método más simple consiste en utilizar el enfoque expuesto al principio del punto 4 del párrafo 3, y en expresar $\alpha_1, \dots, \alpha_r$ a través de variables independientes. En nuestro caso existe una sola variable independiente: supongamos que ésta sea $\alpha_r = \mu$, y $\alpha_A^* = (\mu^*, \dots, \mu^*)$ donde μ^* minimiza

$$|Y - (\mu, \dots, \mu)X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \mu)^2.$$

Es evidente que

$$\mu^* = \bar{y} = \frac{1}{n} \sum_{k=1}^{r} \sum_{i=1}^{n_k} y_{ki},$$

$$|Y - \alpha_A^* X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \bar{y})^2 = Q(Y) =$$
$$= \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \bar{y} + \bar{y}_{k.} - \bar{y}_{k.})^2 =$$
$$= \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \bar{y}_{k.})^2 + \sum_{k=1}^{r} n_k(\bar{y}_{k.} - \bar{y})^2$$

(la suma de los productos mixtos es igual a cero, puesto que $\sum_{i=1} (y_{ki} - \bar{y}_{k.}) = 0$). Si la hipótesis H_1 es cierta, entonces, en virtud de (3.33), (3) y de la igualdad recién obtenida,

$$|(\alpha_A^* - \alpha^*)X|^2 = Q(Y) - Q_2(Y) = \sum_{k=1}^{r} n_k(\bar{y}_{k.} - \bar{y})^2 = Q_1(Y).$$

En virtud de (3.36), al cumplirse H_1 obtenemos $Q_1(Y)/Q_2(Y) \in \mathbf{F}_{r-1,n-r}$, lo cual no da la posibilidad de construir el criterio $Q_1(Y)/Q_2(Y) > f_\varepsilon$ (f_ε es la cuantila de $\mathbf{F}_{r-1,n-r}$ de orden $1 - \varepsilon$) para verificar H_1, el cual será el c.r.v. Si se conoce σ^2, el c.r.v. tendrá la forma

$$Q_1(Y) > \sigma^2 h_\varepsilon$$

(h_ε es la cuantila de H_{r-1}) y será el criterio minimax para las alternativas separadas respectivamente (véase el § 3.9).

2. Influencia de dos factores. Enfoque elemental. En los problemas de este apartado se investiga la influencia que los factores de dos tipos ejercen sobre los resultados del experimento. Con arreglo, digamos, a la agricultura, esto puede ser el estudio de la influencia que ejerce la composición del suelo (el factor A adopta r valores) y el método de cultivo (el factor B adopta s valores) sobre la calidad de la cosecha.

Aquí las observaciones pueden representarse en la forma

$$y_{kli} = \alpha_{kl} + \xi_{kli}, \quad \xi_{kli} \in \Phi_{0,\sigma^2}, \qquad (4)$$
$$k = 1, \ldots, r, \; l = 1, \ldots, s, \; i = 1, \ldots, n_{kl},$$

y el modelo sometido a investigación, en esencia, no se distinguirá en nada del modelo (1) examinado en el punto 1. Por consiguiente, aquí también son aplicables todos los resultados del § 3, pero su aplicación directa es más voluminosa. Ya de por sí es voluminosa la propia presencia de índices triples. Para simplificar algo el problema, pongamos $n_{kl} = 1$; esto nos permitirá eliminar uno de los índices (índice i en (4)). Además, en este apartado proponemos un enfoque elemental algo distinto, que, independientemente de los teoremas del § 3, permitirá obtener las afirmaciones necesarias para la verificación de las hipótesis fundamentales.

Así pues, examinaremos las muestras $Y_{kl} = y_{kl}$ de volumen unitario, de tal modo que el conjunto de datos experimentales Y aquí será la matriz

$r \times s$ de los números y_{kl} que determinan el resultado del experimento bajo la influencia del k-ésimo factor A y el l-ésimo valor del factor B. Esta matriz puede interpretarse como r muestras (filas) de volumen s, correspondientes a distintos valores del factor A, o bien como s muestras (columnas) de volumen r, correspondientes a distintos valores del factor B. De acuerdo con esto, más adelante precisamente tendrá lugar la agrupación de las observaciones. Pongamos

$$\bar{y}_k = \frac{1}{s}\sum_{l=1}^{s} y_{kl}, \quad \bar{y}_{.l} = \frac{1}{r}\sum_{k=1}^{r} y_{kl}, \quad \bar{y} = \frac{1}{rs}\sum_{k,l} y_{kl}.$$

Es válida la identidad

$$Q(Y) \equiv \sum_{k,l} (y_{kl} - \bar{y})^2 = Q_1(Y) + Q_2(Y) + Q_3(Y), \tag{5}$$

donde

$$Q_1(Y) = s\sum_{k} (\bar{y}_{k.} - \bar{y})^2, \quad Q_2(Y) = r\sum_{l} (\bar{y}_{.l} - \bar{y})^2,$$
$$Q_3(Y) = \sum_{k,l} (y_{kl} - \bar{y}_{k.} - \bar{y}_{.l} + \bar{y})^2.$$

Supongamos que la influencia ejercida por los factores es aditiva, o sea, existen a_k y b_l tales que

$$\alpha_{kl} = a_k + b_l, \quad k = 1, \ldots, r, \; l = 1, \ldots, s. \tag{6}$$

Es evidente que Q_1 determina la variabilidad de los valores a_k (o sea, está relacionada con el factor A), Q_2 determina la variabilidad de b_l (factor B), y Q_3 es una suma que se origina absolutamente por casualidad. También es evidente que

$$Q_i(Y + a) = Q_i(Y), \quad i = 1, 2, 3. \tag{7}$$

Teorema 1. 1)

$$Q_3(Y)/\sigma^2 \in \mathbf{H}_{(r-1)(s-1)}. \tag{8}$$

2) *Si es cierta la hipótesis* $H_A = \{a_1 = \ldots = a_r = a\}$, *entonces* $Q_1(Y)$ *no depende de* $Q_2(Y)$ *y* $Q_3(Y)$, $Q_1(Y)/\sigma^2 \in \mathbf{H}_{r-1}$. *Una afirmación análoga tiene lugar respecto a* Q_2, *y la hipótesis* $H_B = \{b_1 = \ldots = b_s = b\}$.

3) *Si es válida la hipótesis* $H_1 = \{\alpha_{kl} = \alpha\}$, todas las formas cuadráticas Q_1, Q_2 y Q_3 son independientes.

Demostración. Pongamos, sin limitar la generalidad, $\sigma^2 = 1$. Entonces

$$\mathbf{M} y_{kl} y_{ij} = \begin{cases} \alpha_{kl}\alpha_{ij}, & \text{si } (i, j) \neq (k, l), \\ \alpha_{kl}^2 + 1, & \text{si } (i, j) = (k, l). \end{cases}$$

De aquí se deduce que

$$\mathbf{M}\left(\sum_{\mathrm{I}} y_{kl}\right)\left(\sum_{\mathrm{II}} y_{kl}\right) = \left(\sum_{\mathrm{I}} \alpha_{kl}\right)\left(\sum_{\mathrm{II}} \alpha_{kl}\right) + m,$$

donde m es el número de sumandos iguales en las sumas $\sum_{\mathrm{I}}$ y $\sum_{\mathrm{II}}$. Utilizando esta igualdad, ahora es fácil obtener que

$$\mathbf{M}(y_{k.} - \bar{y})(\bar{y}_{.l} - \bar{y}) = (\alpha_{k.} - \bar{\alpha})(\alpha_{.l} - \bar{\alpha}) = (a_k - \bar{a})(b_l - \bar{b}) \qquad (9)$$

en caso de acuerdos naturales respecto a las designaciones $\alpha_{k.}$, $\alpha_{.l}$, $\bar{\alpha}$, $\bar{a}$, $\bar{b}$. Si es cierta la hipótesis $H_A = \{a_1 = \ldots = a_r = a\}$, la esperanza matemática en (9) es igual a cero. Como en este caso $\mathbf{M}(\bar{y}_{k.} - \bar{y}) = \alpha_{k.} - \bar{\alpha} = 0$, el hecho establecido quiere decir que el conjunto de variables aleatorias $\{\bar{y}_{k.} - \bar{y}\}$ no depende de $\{\bar{y}_{.l} - \bar{y}\}$.

Análogamente establecemos que para cualesquiera k, l, i,

$$\mathbf{M}(y_{kl} - \bar{y}_{k.})(\bar{y}_{i.} - \bar{y}) = 0.$$

Esto quiere decir que la población $\{\bar{y}_{k.} - \bar{y}\}$ tampoco depende de $\{y_{kl} - \bar{y}_{k.} - \bar{y}_{.l} + \bar{y}\}$. Esto significa, a su vez, que al cumplirse $\mathbf{H}_A$, $Q_1(Y)$ no depende de $Q_2(Y)$ y $Q_3(Y)$. El hecho de que $Q_1(Y) \in \mathbf{H}_{r-1}$, se deduce del lema de Fisher (§ 2.32).

Igualmente sucede cuando se cumple la hipótesis H_B. No obstante, si es válida la hipótesis H_1 (o sea, si son válidas H_A y H_B), es evidente que los tres conjuntos de variables aleatorias mencionadas más arriba serán independientes. Esto significa la independencia de $Q_1(Y)$, $Q_2(Y)$ y $Q_3(Y)$.

Nos queda hallar la distribución $Q_3(Y)$. En vista de que esta distribución no depende de a_k y b_l, podemos considerar que $a_k = b_l = 0$ para todos los k y l y, por consiguiente, se cumple H_1. Entonces, de la definición $Q(Y)$ resulta que $Q(Y) \in \mathbf{H}_{rs-1}$. Además, es válida (5), donde $Q_1(Y) \in \mathbf{H}_{r-1}$ y $Q_2(Y) \in \mathbf{H}_{s-1}$. Nos queda utilizar la independencia $Q_i(Y)$ y el lema 3.1. El teorema está demostrado.

Con arreglo a los problemas del punto 1 también se puede aplicar un enfoque análogo.

Del teorema 1 se deduce la posibilidad de construir los siguientes procedimientos estadísticos:

1) Estimación de los parámetros $a_k - a_l$, $b_l - b_j$, σ^2 (los números a_k y b_l en (6) han sido determinados con una exactitud de hasta el último sumando) con ayuda de las estimaciones $\bar{y}_{k.} - \bar{y}_{i.}$, $\bar{y}_{.l} - \bar{y}_{.j}$, $(\sigma^2)^* = Q_3(Y)/(r-1)(s-1)$. Como, de hecho, las investigaciones realizadas anteriormente coinciden con lo que hemos hecho en el § 3 y en el punto 1 de este párrafo, las estimaciones mencionadas serán eficientes. Los intervalos confidenciales para σ^2, $a_k - a_i$ pueden ser construidos mediante las rela-

ciones (8),

$$\bar{y}_{k.} - \bar{y}_{l.} - (a_k - a_l) \in \Phi_{0,2\sigma^2/s},$$

$$\frac{\bar{y}_{k.} - \bar{y}_{l.} - (a_k - a_l)}{\sqrt{\frac{2Q_3(Y)}{s(r-1)(s-1)}}} \in \mathbf{T}_{(r-1)(s-1)}$$

(para $b_i - b_j$ todo ocurre análogamente).

2) Verificación de la hipótesis H_A con ayuda del criterio $Q_1/Q_3 > f_\varepsilon$. El nivel del criterio constituirá $1-\varepsilon$ si f_ε es una cuantila de orden $1-\varepsilon$ de la distribución $\mathbf{F}_{r-1,(r-1)(s-1)}$.

El criterio para verificar H_B: $Q_2/Q_3 > f_\varepsilon$ tendrá una forma análoga, donde f_ε es una cuantila de orden $1-\varepsilon$ de la distribución $\mathbf{F}_{s-1,(r-1)(s-1)}$.

3) Verificación de la hipótesis H_1 con ayuda del criterio

$$\frac{Q_1 + Q_2}{Q_3} > f_\varepsilon$$

de nivel $1-\varepsilon$, donde f_ε es una cuantila de orden $1-\varepsilon$ de la distribución $\mathbf{F}_{r+s-2,(r-1)(s-1)}$.

Los problemas del análisis de varianza se examinan más detalladamente en [82] y [83].

§ 5. Reconocimiento de imágenes

En este párrafo examinaremos brevemente un grupo de problemas para cuya designación, además del nombre "reconocimiento de imágenes", a veces también se utilizan los términos "clasificación" y "análisis discriminante" *).

En el § 3.1 hemos examinado el siguiente problema de verificación de r hipótesis simples. Se dan las distribuciones $\mathbf{P}_1, \ldots, \mathbf{P}_r$ y la muestra X de volumen n. Es preciso determinar cuál de las hipótesis

$$H_j = \{X \in \mathbf{P}_j\} \tag{1}$$

es cierta.

Sin embargo, en los problemas prácticos, las distribuciones $\mathbf{P}_j$ a menudo se desconocen, y en cuanto a ellas sólo podemos juzgar a partir de las muestras.

Así pues, supongamos que tenemos r muestras $X_l = (x_{l1}, \ldots, x_{ln_l})$, $l = 1, \ldots, r$, de volúmenes $n_1, \ldots, n_r$, respectivamente, que corresponden

*) Cabe señalar que los últimos dos términos también se usan para designar otros problemas, por ejemplo, aquellos en los que se conocen las distribuciones $\mathbf{P}_l$ en (1).

a r distribuciones *desconocidas* $\mathbf{P}_1, \ldots, \mathbf{P}_r$, y supongamos, además, que tenemos la muestra X. Es necesario resolver otra vez el mismo problema: determinar, cuál de las hipótesis (1) es cierta. Con otras palabras, es necesario establecer cuál de las muestras $X_1, \ldots, X_r$ es la prolongación de la muestra X. Este es precisamente el problema de reconocimiento de imágenes.

Para simplificar la exposición nos limitaremos a estudiar el caso de $r = 2$.

1. Caso paramétrico. Al principio supongamos que $\mathbf{P}_i$ pertenece a cierta familia paramétrica $\{\mathbf{P}_\theta\}$ que satisface la condición (A_μ), o sea, $X_1 \in \mathbf{P}_{\theta_1}$, $X_2 \in \mathbf{P}_{\theta_2}$, $X \in \mathbf{P}_\theta$ para ciertos $\theta_1 \neq \theta_2$ y $\theta = \theta_1$ o $\theta = \theta_2$. La primera de estas afirmaciones corresponde a la hipótesis $H_1 = \{X \in \mathbf{P}_{\theta_1}\}$, y la segunda, a la hipótesis $H_2 = \{X \in \mathbf{P}_{\theta_2}\}$.

Supongamos seguidamente, también para simplificar la exposición, que los volúmenes n_1, n_2 y n de las muestras son iguales: $n_1 = n_2 = n$.

Examinemos la muestra unida (X_1, X_2, X) y representémosla como una muestra de volumen n formada por las observaciones (x_{1i}, x_{2i}, x_i) y perteneciente a la distribución $\mathbf{P}_{\theta_1} \times \mathbf{P}_{\theta_2} \times \mathbf{P}_\theta$ que tiene una densidad $f_{\theta_1}(x_1)f_{\theta_2}(x_2)f_\theta(x)$ dependiente del parámetro $\bar{\theta} = (\theta_1, \theta_2, \theta)$. Es evidente que la función de verosimilitud de la muestra (X_1, X_2, X) será igual a

$$f_{\bar{\theta}}(X_1, X_2, X) = f_{\theta_1}(X_1)f_{\theta_2}(X_2)f_\theta(X).$$

Hemos llegado al problema de verificación de la hipótesis H_1 acerca de que el parámetro $\bar{\theta}$ se encuentra en la "curva" $\theta = \theta_1$ frente a la hipótesis alternativa H_2 acerca de que $\bar{\theta}$ se encuentra en otra "curva" $\theta = \theta_2$. Este es el problema de verificación de la hipótesis de pertenencia a una subfamilia paramétrica (véase el § 3.15), pero en el caso cuando la hipótesis alternativa significa la pertenencia a otra subfamilia paramétrica. El examen de este problema es análogo al expuesto en el § 3.15, pero en cuanto a su dificultad técnica sale fuera del marco de este manual. Aquí nos limitaremos a describir brevemente, para el caso del parámetro unidimensional θ, la esencia del resultado. Esta esencia es completamente análoga al contenido del § 3.15: si el parámetro θ ha sido localizado, o sea, si los puntos θ_1 y θ_2 están situados en el entorno de cierto punto θ_0, $|\theta_1 - \theta_2| > b/\sqrt{n}$ y si la familia $\{\mathbf{P}_\theta\}$ satisface en el punto θ_0 las condiciones de regularidad (RR), entonces, el criterio de la relación de verosimilitud

$$\frac{\sup\limits_{\theta_1,\theta_2} f_{\theta_1}(X_1)f_{\theta_2}(X_2)f_{\theta_2}(X)}{\sup\limits_{\theta_1,\theta_2} f_{\theta_1}(X_1)f_{\theta_2}(X_2)f_{\theta_1}(X)} > c \tag{2}$$

será, cuando $n \to \infty$, asintóticamente minimax para verificar H_1 frente a H_2.

La limitación $n_1 = n_2 = n$ no tiene importancia. La misma se elimina al igual que en los planteamientos del § 1.

2. Caso general. En el caso general, cuando no hay razones para suponer que X_i están relacionadas con una familia paramétrica, es posible un enfoque general basado en las mismas ideas que hemos utilizado al construir, en el § 2, los criterios de homogeneidad. En este caso el criterio π para verificar H_1 frente a H_2 será una función de tres muestras, así que $\pi = \pi(X_1, X_2, X)$ será la probabilidad de que se acepte H_2 para (X_1, X_2, X) dadas. Al igual que antes, el criterio no randomizado es definido por la región crítica $\Omega \subset \mathscr{X}^{n_1+n_2+n}$ en el espacio de los valores de (X_1, X_2, X). Por nivel de significación del criterio se entiende el número

$$1 - \varepsilon = \inf_{\mathbf{P}_1 \in \mathscr{P}, \mathbf{P}_2 \in \mathscr{P}} \mathbf{P}_1 \times \mathbf{P}_2 \times \mathbf{P}_1((X_1, X_2, X) \notin \Omega),$$

donde $\mathscr{P}$ es la clase de distribuciones admisibles. El valor

$$\beta_\pi \mathbf{P}_1, \mathbf{P}_2 = \mathbf{P}_1 \times \mathbf{P}_2 \times \mathbf{P}_2((X_1, X_2, X) \in \Omega),$$
$$\mathbf{P}_1 \in \mathscr{P}, \mathbf{P}_2 \in \mathscr{P},$$

es la potencia del criterio en el punto $(\mathbf{P}_1, \mathbf{P}_2)$.

El criterio π se llama conciliable cuando $\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) \to 1$ para $n_1 \to \infty$, $n_2 \to \infty$, $n \to \infty$ y para cualesquiera $\mathbf{P}_1 \neq \mathbf{P}_2$, $\mathbf{P}_1 \in \mathscr{P}$, $\mathbf{P}_2 \in \mathscr{P}$.

Como base para construir los criterios conciliables se puede utilizar el hecho bien conocido, acerca de la aproximación de las distribuciones empíricas $\mathbf{P}^*_{X_1}$ y $\mathbf{P}^*_{X_2}$ para las muestras X_1 y X_2 con $\mathbf{P}_1$ y $\mathbf{P}_2$, respectivamente. Si $d(\mathbf{P}, \mathbf{Q})$ es cierta distancia entre las distribuciones, entonces, en el caso de la hipótesis H_2, la distancia $d(\mathbf{P}^*_{X_2}, \mathbf{P}^*_X)$ debe ser menor que $d(\mathbf{P}^*_{X_1}, \mathbf{P}^*_X)$. Por eso, en calidad de criterio se puede utilizar la desigualdad

$$d(\mathbf{P}^*_{X_2}, \mathbf{P}^*_X) - d(\mathbf{P}^*_{X_1}, \mathbf{P}^*_X) < c$$

que al ser cumplida se acepta H_2. El cálculo de tal tipo de criterios (de sus niveles de significación y de su potencia) suele acompañarse de grandes dificultades (comparadlo con el tipo de problemas más simples dados en el § 2).

Utilizando la *agrupación de observaciones*, en el caso general podemos aplicar el criterio asintóticamente óptimo (2). Supongamos que tal agrupación se ha hecho en las regiones $\Delta_1, \ldots, \Delta_m$ y que $(\nu_{i1}, \ldots, \nu_{im})$ y $(\nu_1, \ldots, \nu_m)$ son las frecuencias con que en estas regiones caen las observaciones de las muestras X_i, $i = 1, 2$, y X, respectivamente. Supongamos, además, que $\theta_i = (\theta_{i1}, \ldots, \theta_{im})$ son las probabilidades $(\mathbf{P}_i(\Delta_i), \ldots, \mathbf{P}_i(\Delta_m))$ de caída en las regiones $\Delta_1, \ldots, \Delta_m$ para las distribuciones $\mathbf{P}_i$, $i = 1, 2$. En vista de que para la muestra agrupada X_i, $i = 1, 2$, la función de verosimilitud

$f_{\theta_i}(X_i)$ es igual a $f_{\theta_i}(X_i) = \prod_{k=1}^{m} \theta_{ik}^{\nu_{ik}}$, el criterio (2) tendrá la forma siguiente:

$$\sup_{\theta_2} \sum_{k=1}^{m} (\nu_{2k} + \nu_k)\ln \theta_{2k} + \sup_{\theta_1} \sum_{k=1}^{m} \nu_{1k} \ln \theta_{1k} -$$
$$- \sup_{\theta_1} \sum_{k=1}^{m} (\nu_{1k} + \nu_k)\ln \theta_{1k} - \sup_{\theta_2} \sum_{k=1}^{m} \nu_{2k} \ln \theta_{2k} > \ln c,$$

o bien

$$\sum_{k=1}^{m} (\nu_{2k} + \nu_k) \ln \frac{\nu_{2k} + \nu_k}{n_2 + n} + \sum_{k=1}^{m} \nu_{1k} \ln \frac{\nu_{1k}}{n_1} >$$
$$> \ln c + \sum_{k=1}^{m} (\nu_{1k} + \nu_k) \ln \frac{\nu_{1k} + \nu_k}{n_1 + n} + \sum_{k=1}^{m} \nu_{2k} \ln \frac{\nu_{2k}}{n_2} . \qquad (3)$$

Los planteamientos análogos también pueden efectuarse para $r > 2$.

CAPÍTULO 5

Enfoque de los problemas de la estadística matemática desde el punto de vista de la teoría de los juegos

En los §§ 1—3 se introducen los conceptos de juegos ordinario y estadístico.

En los §§ 4, 5 se examinan los métodos de búsqueda de las decisiones estadísticas óptimas

El material expuesto en los §§ 6—8 está dedicado a la construcción de las reglas de decisión asintóticamente óptimas.

§ 1. Observaciones preliminares

En los capítulos anteriores hemos examinado una gran cantidad de problemas estadísticos diferentes, unidos, todos ellos, por la circunstancia siguiente: el estadista, basándose en datos experimentales, ha de tomar cierta decisión. En la teoría de las estimaciones, tales decisiones pueden tener forma de estimaciones puntuales θ^*, las cuales deben ser adoptadas en calidad de cierto parámetro desconocido θ. En la teoría de verificación de hipótesis estadísticas, las decisiones pueden adoptar forma de afirmaciones que especifican cuáles suposiciones referentes a la naturaleza del objeto sujeto a investigación son ciertas y cuáles son falsas. Dichas decisiones, al ser erróneas, ofrecen pérdidas ulteriores. Por ejemplo, en la estimación de laboratorio (realizada con la ayuda de una muestra), un error en cuanto al contenido de diversos componentes en el mineral, puede provocar la alteración del régimen óptimo de fusión y el empeoramiento de la calidad del metal fundido. Esto significa que experimentaremos pérdidas materiales, las cuales dependerán de la magnitud del desacierto. Un error relacionado con la eficacia de un medicamento que se comprueba en un grupo elegido de enfermos, evidentemente, también puede provocar pérdidas que, para mante-

ner la uniformidad del enfoque, consideraremos que podrán ser calculadas en ciertas unidades. También tomaremos este mismo acuerdo con respecto a otros problemas de estadística en los que las pérdidas no tienen un carácter material claramente expresado.

Lo dicho nos permite destacar, en los problemas de la estadística matemática, los siguientes cuatro elementos comunes que, de hecho, determinan la esencia de cada problema concreto. Para simplificar la exposición, en lo sucesivo hablaremos exclusivamente de los *problemas de una sola muestra X de volumen fijo n.*

1) Conjunto Θ cuyos elementos $\theta \in \Theta$ determinan el estado del objeto sujeto a investigación. Si se conoce θ no habrá necesidad de construir una decisión estadística. El conjunto Θ también se denomina conjunto de parámetros, aunque θ también pueden admitir una interpretación más amplia (por ejemplo, el conjunto Θ puede ser muy rico y coincidir con el conjunto de todas las distribuciones en cierto espacio $\mathscr{X}$).

2) Para obtener alguna información acerca de θ desconocido, el estadista hace un experimento y realiza observaciones respecto a cierta variable aleatoria cuya distribución depende de θ. Con otras palabras, el estadista dispone de la muestra X de la distribución $\mathbf{P}_\theta$. Como ya sabemos, de dicha muestra se puede extraer la información acerca de $\mathbf{P}_\theta$ y, por consiguiente, acerca de θ. Podemos considerar que se cumple la condición (A_0) (véase el § 2.6) en cuanto a la correspondencia biunívoca entre θ y $\mathbf{P}_\theta$.

3) En los problemas de estadística siempre está determinado el conjunto $D = \{\delta\}$ de decisiones que puede tomar el estadista. En la teoría de estimación, el conjunto D suele coincidir con Θ, pero en los problemas de verificación de hipótesis, el conjunto D es finito y el número de sus elementos equivale a la cantidad de hipótesis que se verifican. Si se conoce θ, la decisión $\delta = \alpha(\theta)$ se determina unívocamente. Si se desconoce θ, la decisión δ ha de ser óptima en cierto sentido. Pero la optimización de las decisiones requiere que tengamos la posibilidad de compararlas. Para esto estimaremos que se ha dado la función de pérdidas que determina cuantitativamente la consecuencia de la toma de decisiones.

4) La función de pérdidas $w(\delta, \theta)$ está definida en $D \times \Theta$ e indica las pérdidas que sufriremos si tomamos la decisión δ, en tanto que el objeto sujeto a investigación, al que se refiere la decisión, se halla en estado θ. Consideraremos que $w(\delta, \theta) > 0$ cuando $\delta \neq \varphi(\theta)$, $w(\varphi(\theta), \theta) = 0$.

Si de los cuatro elementos mencionados retiramos el punto 2) acerca de los datos experimentales, obtendremos el objeto que constituye un *juego ordinario de dos personas*, juego en el que el estadista (investigador) desempeña el papel del primer jugador, y la naturaleza, el papel del segundo jugador.

§ 2. Principales conceptos y teoremas relacionados con el juego de dos personas

1. Juego de dos personas.

Definición 1. Llámase *juego de dos personas* la terna (D, Θ, w) compuesta por los conjuntos D y Θ y por la función w que aplica $D \times \Theta$ en la semirrecta $[0, \infty)$. Los elementos δ del conjunto D se denominan *estrategias* (operaciones) *del jugador* I, los elementos $\theta \in \Theta$ se llaman *estrategias del jugador* II, y w es la *función de pérdidas del jugador* I (o la función de ganancia del segundo jugador) que determina las pérdidas $w(\delta, \theta)$ que sufrirá el jugador I si elige la estrategia δ, y las pérdidas que sufrirá el jugador II si elige la estrategia θ.

El principal objetivo de la teoría de los juegos de dos personas consiste en elegir la estrategia óptima del jugador I que a menudo identificaremos con nosotros. Para esto es necesario ordenar de algún modo el conjunto de estrategias. No es fácil hacerlo, ya que las pérdidas $w(\delta, \theta)$, con cuya ayuda debemos realizar la ordenación, dependen de dos argumentos, así que, para cada θ, la estrategia δ que minimiza $w(\delta, \theta)$ será, hablando en general, su propia estrategia.

Definición 2. Diremos que la estrategia δ_1 *es mejor que* δ_2, si

$$w(\delta_1, \theta) \leqslant w(\delta_2, \theta) \quad \textit{para todos } \theta \in \Theta \qquad (1)$$

y si existe por lo menos un valor de $\theta_1 \in \Theta$ para el cual $w(\delta_1, \theta_1) < w(\delta_2, \theta_1)$.

Si sólo se cumple (1), diremos que la estrategia δ_1 *no es peor que* δ_2.

La estrategia δ_0 para la cual

$$w(\delta_0, \theta) \leqslant w(\delta, \theta) \quad \textit{para todos } \delta \text{ y } \theta$$

la llamaremos estrategia *uniformemente óptima* (o uniformemente mejor).

La estrategia uniformemente mejor asegura las pérdidas mínimas para todos θ. No obstante, por regla general, tales estrategias no existen.

Señalaremos los tres enfoques siguientes para investigar las estrategias óptimas del jugador I:

— determinación de las estrategias uniformemente óptimas en las subclases;

— determinación de las estrategias bayesianas y minimax;

— estudio de la población de todas las estrategias no mejorables (de la llamada clase completa de estrategias).

2. Estrategias uniformemente óptimas en las subclases. Con arreglo a los problemas de la estadística matemática se utiliza a menudo el procedimiento siguiente (véase el § 5). De algunas consideraciones no relacionadas directamente con las pérdidas (consideraciones de simetría, naturalidad del procedimiento, simplicidad de los cálculos, etc.) a veces es posible reducir la clase de estrategias sujetas a examen. Si esta reducción es tal que después

de ella existe una estrategia uniformemente óptima, entonces, asimismo se resuelve el problema de elección de la estrategia. Este enfoque debe ir acompañado de investigaciones de la cuestión acerca de si hemos perdido o no (tras reducir la clase) la posibilidad de obtener un resultado mucho mejor. Ejemplos de utilización de tal enfoque (aunque referentes a un objeto más complejo: a los juegos estadísticos) serán examinados en los dos párrafos siguientes. El lector ya sabe de ellos por los capítulos 2 y 3 donde hemos examinado las mejores estimaciones (eficaces) en la subclase de estimaciones no desplazadas, así como los criterios uniformemente más potentes en las subclases de todos los criterios invariantes o no desplazados.

3. Estrategias bayesianas. Estas surgen en los casos en que el segundo jugador elige su estrategia al azar, con cierta distribución (conocida o desconocida) en Θ.

Para tener la posibilidad de examinar posteriormente las estrategias "aleatorias", vamos a suponer que en Θ y D están separadas ciertas σ-álgebras naturales de los subconjuntos $\mathscr{F}_\Theta$ y $\mathscr{F}_D$. Entonces, en $(\Theta, \mathscr{F}_\Theta)$ y $(D, \mathscr{F}_D)$ se pueden definir las distribuciones $\mathbf{Q}$ y π, respectivamente, así que $(\Theta, \mathscr{F}_\Theta, \mathbf{Q})$ y $(D, \mathscr{F}_D, \pi)$ serán los espacios probabilísticos.

La designación de las distribuciones π y $\mathbf{Q}$ induce el espacio probabilístico $(D \times \Theta, \mathscr{F}_{D\times\Theta}, \pi \times \mathbf{Q})$, donde $\mathscr{F}_{D\times\Theta}$ es la σ-álgebra engendrada por los productos directos de los conjuntos de $\mathscr{F}_D$ y $\mathscr{F}_\Theta$. La elección de las σ-álgebras de $\mathscr{F}_D$ y $\mathscr{F}_\Theta$ debe ser tal, que se cumplan las dos condiciones siguientes:

a) $\mathscr{F}_D$ y $\mathscr{F}$ *contienen los conjuntos unipuntuales* $\{\delta\}$ y $\{\theta\}$.

b) La función de pérdidas $w(\delta, \theta)$ es medible con respecto a $\mathscr{F}_{D\times\Theta}$.

Definición 3. Las distribuciones π en $(D, \mathscr{F}_D)$ y $\mathbf{Q}$ en $(\Theta, \mathscr{F}_\Theta)$ se llamarán *estrategias mixtas* o *randomizadas* de los jugadores I y II, respectivamente.

La distribución $\mathbf{Q}$ será frecuentemente llamada distribución *a priori*. El sentido de este término debe estar claro de los capítulos 2 y 3. Además, lo aclararemos adicionalmente en el párrafo siguiente. Los conjuntos de todas las estrategias mixtas de los jugadores I y II (o sea, los conjuntos de todas las distribuciones en $(D, \mathscr{F}_D)$ y $(\Theta, \mathscr{F}_\Theta)$ serán designados por $\tilde{D}$ y $\tilde{\Theta}$. En vista de que $\mathscr{F}_D$ y $\mathscr{F}$ contienen conjuntos unipuntuales, entonces $\tilde{D}$ y $\tilde{\Theta}$ contendrán las distribuciones concentradas en un punto y, por consiguiente, podemos considerar que $\tilde{D}$ y $\tilde{\Theta}$ contienen las estrategias δ y θ que llamaremos *estrategias puras*, a fin de tener la posibilidad de separarlas. El acuerdo, según el cual designaremos con los mismos símbolos δ y θ, respectivamente, las distribuciones de $\tilde{D}$ y $\tilde{\Theta}$ concentradas en un mismo punto δ o θ, no provocará equivocaciones de ningún tipo.

Ahora, las pérdidas $\tilde{w}(\pi, \mathbf{Q})$ provocadas por el uso de estrategias mixtas serán definidas por la igualdad

$$\tilde{w}(\pi, \mathbf{Q}) = \mathbf{M}_{\pi\times Q} w(\delta, \theta) = \int w(u, t)\pi(du)\mathbf{Q}(dt). \qquad (2)$$

Así pues, a la par con el juego inicial podemos examinar el juego, $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ con la función de pérdidas (2), el cual se llama *promediación* o *randomización del juego* (D, Θ, w).

Según el acuerdo adoptado escribiremos

$$\tilde{w}(\pi_{(\delta)}, Q) = \tilde{w}(\delta, Q), \ \tilde{w}(\pi, Q_{(\theta)}) = \tilde{w}(\pi, \theta),$$
$$\tilde{w}(\delta, \theta) = w(\delta, \theta),$$

si $\pi_{(\delta)}$ y $Q_{(\delta)}$ son distribuciones concentradas en los puntos δ y θ, respectivamente.

Es evidente que la randomización del juego (D, Θ, w) significará el paso a un juego con conjuntos de estrategias más ricas, respecto al cual el par inicial es un juego "insertado" que se obtiene al examinar exclusivamente las estrategias puras de ambos jugadores. Como veremos más adelante, los problemas de ordenación de las estrategias en los juegos (D, Θ, w) y $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ se hallan íntimamente ligados.

Definición 4. La estrategia $\pi = \pi_Q$, para la cual

$$\tilde{w}(\pi_Q, Q) = \inf_{\pi} \tilde{w}(\pi, Q),$$

se denomina *estrategia bayesiana, correspondiente a la distribución a priori* Q.

Así pues, la estrategia bayesiana no es otra cosa sino la mejor estrategia π para Q dada en un juego promediado.

La estrategia $\delta_Q \in D$, para la cual $\tilde{w}(\delta_Q, Q) = \inf \tilde{w}(\pi, Q)$, se denomina *estrategia bayesiana pura.*

Teorema 1. *Si para* Q *dada existe una estrategia bayesiana mixta* π_Q, *entonces también existirá una estrategia bayesiana pura* δ_Q *tal, que*

$$\tilde{w}(\delta_Q, Q) = \tilde{w}(\pi_Q, Q).$$

La **demostración** es casi evidente. Designemos $a = \tilde{w}(\pi_Q, Q)$. Está claro que

$$\tilde{w}(\delta, Q) \geqslant \inf_{\delta} \tilde{w}(\delta, Q) \geqslant a.$$

Si admitimos que $\tilde{w}(\delta, Q) > a$ para todas δ, entonces, realizando la mediación respecto a δ con ayuda de π_Q, obtenemos

$$a = \int \tilde{w}(u, Q)\pi_Q(du) > a.$$

Esta contradicción demuestra el teorema. ◁

Ahora bien, si se alcanza $\inf_{\pi} \tilde{w}(\pi, Q)$, esto también se alcanzará en las estrategias puras.

Si no se alcanza $\inf_{\delta} \tilde{w}(\delta, Q)$, entonces no existirán estrategias bayesianas. En este caso resulta útil el concepto de *estrategia ε-bayesiana* que existe

siempre y la cual se define como una estrategia δ_Q para la cual

$$\tilde{w}(\delta_Q, Q) \leqslant \inf_{\delta} \tilde{w}(\delta, Q) + \varepsilon \qquad (3)$$

para $\varepsilon > 0$ dado. Sin embargo, en lo sucesivo, para simplificar la exposición nos limitaremos a examinar tan sólo los problemas que contienen las estrategias bayesianas.

La cuestión acerca de la utilización práctica de las estrategias bayesianas es bastante delicada. Si la existencia de la distribución a priori se debe a cierto mecanismo físico real, este enfoque será indiscutible. Pero el enfoque bayesiano también puede ser justificado en los casos en que el mismo esté relacionado con la existencia de ciertas ideas, quizás subjetivas y no siempre bastante completas, las cuales, no obstante, no deben ser rechazadas. En el apartado siguiente (punto 4) se ofrece un análisis más detallado del asunto relacionado con la utilización del enfoque bayesiano.

4. Estrategias minimax. Si se carece de una información a priori respecto a θ, al ordenar las estrategias es posible orientarse hacia la "peor" estrategia del adversario. Si eligemos la estrategia δ, las pérdidas máximas constituirán

$$\sup_{\theta} w(\delta, \theta) \equiv w(\delta, \uparrow). \qquad (4)$$

Esta cantidad sólo depende de δ y, al igual que los valores de $w(\delta, Q)$, permite ordenar δ.

Definición 5. La estrategia $\bar{\delta}$ se llama *minimax* si

$$w(\bar{\delta}, \uparrow) = \inf_{\delta} w(\delta, \uparrow) \equiv w^*. \qquad (5)$$

El término minimax se forma a base de la unión de las denominaciones de las operaciones en el segundo miembro de la relación

$$w(\bar{\delta}, \uparrow) = \min_{\delta} \max_{\theta} w(\delta, \theta).$$

Es evidente que las estrategias minimax, al igual que las bayesianas, pueden, hablando en general, no existir. En este caso, de un modo análogo a (3), se puede introducir el concepto de estrategia ε-minimax. En los planteamientos ulteriores partiremos del hecho de que en (4) y (5) se alcanzan sup e inf.

En vista de que para cualquier θ

$$w(\bar{\delta}, \theta) \leqslant w(\bar{\delta}, \uparrow) = w^*,$$

la estrategia minimax $\bar{\delta}$ se caracteriza por el hecho de que asegura las pérdidas del jugador 1 en cantidad no mayor de w^*.

Definición 6. Los valores

$$w^* = \inf_{\delta} w(\delta, \uparrow) \quad (w(\delta, \uparrow) = \sup_{\theta} w(\delta, \theta)),$$

$$w_* = \sup_{\theta} w(\downarrow, \theta) \quad (w(\downarrow, \theta) = \inf_{\delta} w(\delta, \theta))$$

se llaman, respectivamente, precio superior e inferior del juego. Si $w^* = w_*$, se dice que existe el precio del juego, igual al valor común de w^* y w_*.

De lo dicho anteriormente y de las consideraciones de simetría está claro que el jugador II, actuando análogamente al primero y eligiendo su estrategia $\bar{\theta}$ de las mismas consideraciones minimax, siempre puede asegurar para sí una ganancia no menor de w^*. (Tal estrategia $\bar{\theta}$ sería más correcto llamarla estrategia maximin, pero para ella utilizaremos el mismo término: estrategia minimax). Por lo tanto, *si existe precio del juego*, entonces, eligiendo la estrategia minimax $\bar{\delta}$, aseguraremos para nosotros un *resultado inmejorable* desde el punto de vista siguiente: si el adversario elige $\bar{\theta}$, ninguna otra estrategia nos causará pérdidas θ menores de $w_* = w^*$. Es evidente que

$$w(\bar{\delta}, \bar{\theta}) = w^* = w_*.$$

En el caso general siempre $w^* \geqslant w_*$, ya que para todos δ y θ $w(\delta, \uparrow) \geqslant w(\delta, \theta) \geqslant w(\downarrow, \theta)$ y, por consiguiente,

$$w^* = \inf_\delta w(\delta, \uparrow) \geqslant \sup_\theta w(\downarrow, \theta) = w_*. \tag{6}$$

Si $w^* > w_*$, entonces, la estrategia minimax $\bar{\delta}$ *se puede mejorar* introduciendo las estrategias mixtas. En esto consiste una de las finalidades principales de estas últimas.

Las estrategias minimax para un juego promediado (si ellas existen) las designaremos por $\bar{\pi}$ y $\bar{Q}$, respectivamente, y pongamos

$$\tilde{w}^* = \inf_\pi \sup_Q \tilde{w}(\pi, Q), \quad \tilde{w}^* = \sup_Q \inf_\pi \tilde{w}(\pi, Q).$$

Mostremos primeramente que, al promediar el juego, los precios superior e inferior de éste se aproximan.

Teorema 2. $w^* \geqslant \tilde{w}^* \geqslant \tilde{w}_* \geqslant \tilde{w}_*$.

La demostración de este teorema, al igual que la del teorema 1, es muy fácil. En vista de que la mediación del juego puede realizarse en dos etapas: primero por el conjunto D y luego por Θ, para la demostración es suficiente examinar tan sólo la promediación parcial $(\tilde{D}, \Theta, \tilde{w})$ del juego (D, Θ, w). Tenemos

$$\tilde{w}^* = \inf_\pi \sup_\theta \tilde{w}(\pi, \theta) \leqslant \inf_\delta \sup_\theta w(\delta, \theta) = w^*.$$

Como para todos π,

$$\tilde{w}(\pi, \theta) = \int w(u, \theta)\pi(du) \geqslant \inf_\delta w(\delta, \theta) = w(\downarrow, \theta),$$

entonces, $\inf \tilde{w}(\pi, \theta) \geqslant w(\downarrow, \theta)$,

$$\tilde{w}^* = \sup_\theta \inf_\pi \tilde{w}(\pi, \theta) \geqslant \sup_\theta w(\downarrow, \theta) = w_*.$$

La desigualdad $\tilde{w}^* \geqslant \tilde{w}_*$ ha sido demostrada en (6). ◁

El hecho fundamental de la teoría de los juegos consiste en el llamado teorema del minimax, el cual afirma que para suposiciones muy amplias, los *juegos promediados tienen un precio de* $\tilde{w}^* = \tilde{w}_*$ *y para ellos existen estrategias minimax.*

Esta afirmación será enunciada más exactamente en el párrafo siguiente, en una situación más general, con arreglo a los juegos estadísticos.

El juego inicial (D, Θ, w), sobre todo en el caso cuando D y Θ son finitos, por regla general no tiene precio.

Ejemplo 1. Examinemos un juego elemental cuando los conjuntos D y Θ son bipuntuales, $D = \{\delta_1, \delta_2\}$, $\Theta = \{\theta_1, \theta_2\}$. Los valores de la función de pérdidas $w(\delta, \theta)$ se definen por la matriz $\| w(\delta_i, \theta_j) \|$, $i, j = 1, 2$, la cual supondremos que es igual a $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$. Esto corresponde, por ejemplo, al juego de adivinación, cuando el jugador I debe adivinar en qué mano el jugador II ha escondido una moneda. La adivinación significa una pérdida nula $(w(\delta_1, \theta_1) = w(\delta_2, \theta_2) = 0)$, y el error, una pérdida igual a 1 rublo $(w(\delta_1, \theta_2) = w(\delta_2, \theta_1) = 1)$. Es evidente que aquí $w(\delta_i, \uparrow) = 1$, $w^* = 1$, $w(\downarrow, \theta_i) = 0$, $w_* = 0$, por consiguiente, el juego no tiene precio, y el jugador I no puede garantizar para sí una pérdida inferior a 1 rublo. El propio concepto de estrategia minimax aquí es inútil.

Examinemos ahora la promediación de este juego. Aquí las clases de estrategias $\tilde{D}$ y Θ son la población de todas las distribuciones en un conjuto bipuntual. Es evidente que cada una de las distribuciones en D y Θ se describe por una probabilidad p y q de elegir las estrategias δ_1 y θ_1, respectivamente. Por eso se puede considerar que $\tilde{D} = [0, 1]$, $\tilde{\Theta} = [0, 1]$. Las pérdidas del jugador I en este juego son iguales a

$$\tilde{w}(p, q) = p(1 - q) + q(1 - p) = p + q - 2pq,$$
$$\tilde{w}(p, \uparrow) = \begin{cases} p + 1 - 2p = 1 - p & \text{para} \quad 2p < 1, \\ p & \text{para} \quad 2p \geqslant 1, \end{cases}$$
$$\tilde{w}^* = 1/2.$$

De un modo análogo hallamos que $\tilde{w}_* = 1/2$. Ahora bien, el juego promediado ya tiene precio y el primer jugador, eligiendo δ_1 y δ_2 con probabilidad $p = 1 - p = 1/2$, puede garantizar para sí una pérdida no mayor de 1/2. Esta estrategia no puede ser mejorada, ya que el jugador II puede garantizar para sí esa misma ganancia, eligiendo $q = 1/2$.

Pero si resulta que el juego promediado no tiene precio (lo cual puede tener lugar tan sólo en los juegos de estructura compleja especial), entonces, la promediación reiterada no dará ningunos resultados, ya que esta promediación repetida coincidirá, en esencia, con la promediación ordinaria.

Los enfoques bayesiano y minimax de la resolución de los problemas de juego tienen gran aplicación en la actividad humana cotidiana. El enfoque bayesiano está orientado hacia la existencia de ciertas nociones, aunque sean aproximadas, del comportamiento del segundo jugador. El enfoque minimax está justificado en los casos en que debemos asegurarnos de una gran derrota.

Ejemplo 2. Un estudiante se prepara para el examen. Supongamos que no es un estudiante ideal y que no ha tenido tiempo suficiente para repasar bien todo el material. Además, el objetivo de este estudiante consiste en obtener la mejor nota posible.

En las condiciones descritas, el estudiante sólo puede estudiar perfectamente parte del material. Por eso, para él son posibles por lo menos dos vías: 1) estudiar en sobresaliente tan sólo las partes que, según la información disponible, el examinador pregunta con más frecuencia; 2) estudiar un poco todo el mateiral para asegurarse una nota buena o satisfactoria. La primera variante corresponderá al enfoque bayesiano, y la segunda, al enfoque minimax.

Claro está que la estrategia uniformemente óptima aquí sería estudiar perfectamente todo el material, pero, según la condición del problema, tal estrategia no es posible.

En las situaciones concretas, las estrategias minimax no siempre son racionales.

Ejemplo 3. Supongamos que $\Theta = [0, 1]$ y que el conjunto $D = \{\delta_1, \delta_2\}$ consta de dos elementos. La función de pérdidas se define por las relaciones (fig. 9)

$$w(\delta_1, \theta) = 1,$$
$$w(\delta_1, \theta) = 4(1 + \varepsilon)\theta(1 - \theta).$$

Fig. 9

Aquí $w(\delta_1, \dagger) = 1$, $w(\delta_1, \dagger) = 1 + \varepsilon$, $w^* = 1$, y δ_1 será la estrategia minimax, aunque en caso de $\varepsilon > 0$ pequeños, para la "mayoría" de los valores de θ, la estrategia δ_2 será mejor: $w(\delta_2, \theta) < 1$ para θ de la región $\left|\theta - \frac{1}{2}\right| > \frac{1}{2}\sqrt{\frac{\varepsilon}{1 + \varepsilon}}$. Para la "mayoría" de las distribuciones Q en

$\Theta = [0, 1]$ (cuya masa no está concentrada en el entorno del punto $\theta = 1/2$), las estrategias bayesianas también coincidirán con δ_2.

Los conceptos de estrategia bayesiana y minimax están relacionados entre sí. La siguiente afirmación proporciona el método de averiguación de las estrategias minimax con ayuda de las estrategias bayesianas.

Definición 7. La estrategia $\bar{\pi}$ se llama *igualadora* en el conjunto $\Theta_0 \subset \Theta$ si

1) $\tilde{w}(\bar{\pi}, \theta) = c = \text{const}, \ \theta \in \Theta_0$,
2) $\tilde{w}(\bar{\pi}, \theta) \leqslant c$ para todos θ.

Teorema 3. *Supongamos que existe la distribución a priori $\bar{\mathbf{Q}}$ y su estrategia bayesiana correspondiente $\pi_{\bar{Q}}$, la cual es igualadora en el portador $N_{\bar{Q}}$ de la distribución $\bar{\mathbf{Q}}$. Entonces, $\bar{\pi} = \pi_{\bar{Q}}$ es una estrategia minimax.*

Si $N_{\bar{Q}} = \Theta$, la estrategia igualadora $\bar{\pi}$ hace "indiferente" el juego del segundo jugador, o sea, lo hace independiente de éste (compárese con el ejemplo 1).

Demostración del teorema 3. Designemos $\sup\limits_{\theta} \tilde{w}(\pi, \theta) = \tilde{w}(\pi, \uparrow)$, $\inf\limits_{\delta} \tilde{w}(\delta, \mathbf{Q}) = \tilde{w}(\downarrow, \mathbf{Q})$. Debemos convencernos de que

$$\tilde{w}(\pi_{\bar{Q}}, \uparrow) = \inf_{\pi} \tilde{w}(\pi, \uparrow).$$

Esto se deduce de las desigualdades siguientes, válidas para cualquier π:

$$\tilde{w}(\pi, \uparrow) \geqslant \tilde{w}(\pi, \bar{\mathbf{Q}}) \geqslant \tilde{w}(\pi_{\bar{Q}}, \bar{\mathbf{Q}}) =$$
$$= \int \tilde{w}(\pi_{\bar{Q}}, t)\bar{\mathbf{Q}}(dt) = c \geqslant \tilde{w}(\pi_{\bar{Q}}, \uparrow). \ \triangleleft$$

A veces es útil la siguiente pequeña generalización del teorema 3.

Teorema 3A. *Supongamos que existen tales sucesiones $\mathbf{Q}_n$, π_{Q_n} que $\tilde{w}(\pi_{Q_n}, \mathbf{Q}_n) \to c$. Supongamos, además, que existe una estrategia $\bar{\pi}$ dotada de la propiedad $w(\bar{\pi}, \theta) \leqslant c$ para todos θ. Entonces, $\bar{\pi}$ es la estrategia minimax.*

La **demostración** es igualmente fácil:

$$\tilde{w}(\pi, \uparrow) \to \tilde{w}(\pi, \mathbf{Q}_n) \geqslant \tilde{w}(\pi_{Q_n}, \mathbf{Q}_n) \to c.$$

Esto puede tener lugar si y sólo si $\inf\limits_{\pi} \tilde{w}(\pi, \uparrow) \geqslant c$. Como $c \geqslant w(\bar{\pi}, \uparrow)$, el teorema queda demostrado.

La distribución $\bar{\mathbf{Q}}$ en el teorema 3, que define la estrategia minimax bayesiana $\pi_{\bar{Q}}$, posee una propiedad magnífica: la misma será la peor en el sentido de que las pérdidas bayesianas $\tilde{w}(\pi_Q, q)$ serán máximas para ella.

Definición 8. La distribución $\bar{\mathbf{Q}}$ se denomina *la menos favorable* o *la peor*, si

$$\tilde{w}(\pi_{\bar{Q}}, \bar{\mathbf{Q}}) = \sup_{Q} \tilde{w}(\pi_Q, \mathbf{Q}),$$

o, con otras palabras, $\tilde{w}(\downarrow, \bar{\mathbf{Q}}) = \sup\limits_{Q} \tilde{w}(\downarrow, \mathbf{Q})$.

Teorema 4. *Supongamos que el juego* $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ *tiene precio y que ambos jugadores tienen estrategias minimax* $\bar{\pi}$ y $\bar{Q}$. *Entonces, la distribución* $\bar{Q}$ *es la peor,* y $\bar{\pi}$ es la estrategia bayesiana $\bar{\pi} = \pi_{\bar{Q}}$ que responde a $\bar{Q}$.

Observación 1. Del hecho de que, en virtud del teorema 1, a la par con $\pi_{\bar{Q}}$ existe la estrategia bayesiana pura $\delta_{\bar{Q}}$, de ningún modo se deduce que esta última también será minimax.

Observación 2. En virtud del teorema fundamental de los minimax, la condición del teorema 4 acerca de la existencia de precio del juego promediado y de estrategias minimax, no se debe considerar como una limitación considerable.

Necesitaremos la siguiente afirmación auxiliar que enunciaremos en términos del juego inicial (no promediado).

Lema 1. *Supongamos que el juego* (D, Θ, w) *tiene precio y estrategias minimax* $\bar{\delta}$ *y* $\bar{\theta}$ *de ambos jugadores:*

$$w(\bar{\delta}, \uparrow) = \inf_{\delta} w(\delta, \uparrow), \quad w(\downarrow, \bar{\theta}) = \sup_{\theta} w(\downarrow, \theta).$$

Entonces

$$w(\bar{\delta}, \uparrow) = w(\bar{\delta}, \bar{\theta}) = w(\downarrow, \bar{\theta}), \tag{7}$$

$$w^* = w(\bar{\delta}, \bar{\theta}) = w_*. \tag{8}$$

Al contrario, si para ciertos $\bar{\delta}, \bar{\theta}$ *se cumple* (7), *entonces es válida* (8), y $\bar{\delta}, \bar{\theta}$ *son estrategias minimax.*

Demostración. Para todos δ y θ tenemos

$$w(\delta, \uparrow) \geqslant w(\delta, \theta) \geqslant w(\uparrow, \theta).$$

De aquí resulta

$$w^* = w(\bar{\delta}, \uparrow) \geqslant w(\bar{\delta}, \bar{\theta}) \geqslant w(\downarrow, \bar{\theta}) = w_*. \tag{9}$$

Como, según la condición, $w^* = w_*$, en (9) todos los signos de desigualdad deben sustituirse por signos de igualdad. Esto demuestra (7) y (8).

Al contrario, si es válida (7), entonces

$$w^* = \inf_{\delta} w(\delta, \uparrow) \leqslant w(\bar{\delta}, \uparrow) = w(\downarrow, \bar{\theta}) \leqslant \sup_{\theta} w(\downarrow, \theta) = w_*.$$

En vista de que siempre $w^* \geqslant w_*$, las desigualdades mencionadas significan que $w^* = w_*$ y que las estrategias $\bar{\delta}$ y $\bar{\theta}$ son minimax. El lema queda demostrado.

El punto $(\bar{\delta}, \bar{\theta})$ que posee la propiedad (7) se llama *punto de ensilladura*, el lema 1 se denomina criterio de existencia del punto de ensilladura de las estrategias minimax inmejorables.

Demostración del teorema 4. Apliquemos el lema 1 al juego promediado $(\tilde{D}, \tilde{\Theta}, \tilde{w})$. Entonces obtendremos que

$$\tilde{w}(\bar{\pi}, \bar{Q}) = \tilde{w}(\downarrow, \bar{Q}) = \tilde{w}_* = \sup_{Q} \tilde{w}(\downarrow, Q).$$

De aquí se desprende que la distribución $\bar{Q}$ es la peor y que $\bar{\pi}$ es la estrategia bayesiana correspondiente a $\bar{Q}$. El teorema queda demostrado.

El contenido de las afirmaciones citadas anteriormente ahora se puede resumir en forma del criterio siguiente, que tiene carácter minimax y que describe muy ampliamente la relación entre las estrategias minimax y las estrategias bayesianas.

Teorema 5. *Supongamos que el juego* ($\tilde{D}$, Θ, $\tilde{w}$) *tiene precio y estrategias minimax. Entonces, las tres condiciones siguientes son equivalentes:*

1) *La estrategia* $\bar{\pi}$ *es minimax.*

2) *La estrategia* $\bar{\pi}$ *es bayesiana e igualadora.*

3) *La estrategia* $\bar{\pi}$ *es bayesiana y corresponde a la peor distribución* $\bar{Q}$: $\bar{\pi} = \pi_{\bar{Q}}$.

Demostración. La relación 2) ⇒ 1) se ha demostrado en el teorema 3 (para esto no se necesita la condición del teorema 5). La relación 1) ⇒ 3) se ha establecido en el teorema 4. Necesitamos convencernos de que 3) ⇒ 2), o sea, que la estrategia bayesiana, correspondiente a la peor distribución, es igualadora. Tenemos

$$\tilde{w}_* = \tilde{w}(\bar{\pi}, \bar{Q}) = \int \tilde{w}(\bar{\pi}, t)\bar{Q}(dt) \leqslant \sup_t \tilde{w}(\bar{\pi}, t) = \tilde{w}^*.$$

Esto significa que $\int \tilde{w}(\bar{\pi}, t)\bar{Q}(dt) = \sup_t \tilde{w}(\pi, t)$ y, por consiguiente,

$$\tilde{w}(\bar{\pi}, t) = \tilde{w}(\bar{\pi}, \dagger) \text{ c.d. } [\bar{Q}].$$

En vista de que, además, siempre $\tilde{w}(\bar{\pi}, t) \leqslant \tilde{w}(\bar{\pi}, \dagger)$, entonces $\bar{\pi}$ es una estrategia ingualadora. El teorema queda demostrado.

Volvamos ahora a la cuestión acerca de la aplicación de las clases examinadas de estrategias. Supongamos que no podemos destacar la subclase de estrategias que nos satisfagan, entre las cuales exista la estrategia uniformemente mejor. Supongamos, seguidamente, que disponemos de ciertas nociones acerca del comportamiento del segundo jugador (o sea, de los valores estimados de θ) que, sin embargo, no son suficientes para aplicar el enfoque bayesiano en su forma pura. En estas condiciones el enfoque minimax significará el desprecio de la información que tenemos a nuestra disposición. En tal situación se puede utilizar el enfoque intermedio que consiste en lo siguiente:

1) Primero es necesario protegerse contra las altas pérdidas, o sea, examinar tan sólo las estrategias δ para las cuales $w(\delta, \theta) \leqslant w^* + a$ con valores convenientes de $a > 0$ y para todos θ. El conjunto de estrategias que satisfacen esta desigualdad serán designadas por D_a.

2) En este subconjunto (o sea, en el juego (D_a, Θ, w) ya se puede aplicar el enfoque bayesiano, utilizando las aproximaciones, accesibles a nosotros, para la distribución a priori Q.

Tal enfoque mixto se usa también constantemente en la actividad humana cotidiana. En las condiciones del ejemplo 2 este enfoque significará que el estudiante aprenderá muy superficialmente todo el material (para evitar una nota insatisfactoria) y luego aprenderá mejor lo que se pregunta con más frecuencia.

La utilización matemática del enfoque mixto debe acompañarse de investigaciones de la estabilidad de las pérdidas bayesianas en el juego (D_0, Θ, w) para las variaciones admisibles de Q.

5. Clase completa de estrategias. Si todos los enfoques anteriormente descritos no permiten elegir unívocamente la estrategia, la solución del problema se limite a la descripción de la llamada clase completa de estrategias.

Definición 9. La clase de estrategias $D^c \subset \tilde{D}$ se llama *completa* si para todo $\pi \notin D^c$ existe la estrategia $\pi_0 \in D^c$ que es mejor que π.

La clase D_0^c se denomina *clase completa mínima* si D_0^c es una clase completa, pero a condición de que ninguna de sus propias subclases no sea una clase completa.

Con otras palabras, la clase completa mínima se compone únicamente de estrategias inmejorables.

La utilidad de construcción de la clase completa mínima o de la clase completa, la cual es mucho menor que D, es evidente. Esto da la posibilidad de reducir el juego ($\tilde{D}$, $\tilde{\Theta}$, $\tilde{w}$) al (D^c, $\tilde{\Theta}$, $\tilde{w}$), el cual puede tener una estructura más simple.

El segundo teorema fundamental de la teoría de los juegos consiste en que para amplias suposiciones, *la clase de todas las estrategias bayesianas* $\{\pi_Q\}$, $Q \in \tilde{\Theta}$, *es una clase completa.* La enunciación exacta de este teorema se dará en el párrafo siguiente. En algunos casos, las clases completas se pueden construir también directamente, utilizando la estructura del juego. Admitamos, por ejemplo, que existe una partición del espacio D en subconjuntos D_b, $D = \bigcup_{b \in B} D_b$, $D_{b_1} \neq D_{b_2}$ cuando $b_1 = b_2$, tal que en cada uno de estos subconjuntos (o sea, para los juegos (D_b, Θ, w)) existe la estrategia uniformemente óptima $\delta_b \in D_b$. Está claro que en este caso la clase $D^c = \{\delta_b\}_{b \in B}$ será completa. Tal enfoque de la construcción de la clase completa será ilustrado en el § 3.

§ 3. Juegos estadísticos

1. Descripción de los juegos estadísticos. Los elementos principales del juego estadístico se forman por la misma terna (D, Θ, w) que hemos examinado en el párrafo precedente. No obstante, se les añade lo siguiente:

1) En los juegos estadísticos el estadista (investigador) desempeña el papel del jugador I, y la naturaleza (más exactamente, la naturaleza del fenó-

meno que se investiga), el papel del jugador II. La naturaleza elige (o "adivina") el parámetro (estrategia) θ que desconocemos y que determina el estado del objeto sometido a investigación. La mayoría de los problemas de la estadística matemática está relacionada, de un modo u otro, con la toma de tales decisiones δ que adivinarían lo más precisamente posible este θ desconocido. En este caso es necesario tener presente que la naturaleza como jugador no tiene por objeto la ganancia máxima (es decir, no intenta causarnos las pérdidas máximas) y desde este punto de vista es un jugador "imparcial" de la elección de sus propias estrategias.

2) En los juegos estadísticos tenemos la posibilidad de "explorar" la estrategia de la naturaleza con ayuda de los experimentos que nos dan en forma de la muestra $X \in \mathbf{P}_\theta$ las indicaciones "sugestivas" de cuál debe ser el valor de θ. Pues, la muestra X de volumen *) n, procedente de la distribución $\mathbf{P}_\theta$ que depende de θ, es un elemento del juego estadístico.

En estas condiciones debemos elegir, evidentemente, nuestra decisión δ en dependencia de X. Por consiguiente, ahora llegan a ser estrategias del estadista todas las funciones $\delta(X)$ que aplican $\mathscr{X}^n$ en D. Estas funciones $\delta(X)$ se llaman *funciones de decisión* o *reglas de decisión*. Nos limitaremos a examinar sólo las funciones $\delta(X)$ que realizan la aplicación medible de $(\mathscr{X}^n, \mathfrak{B}^n_{\mathscr{X}})$ en $(D, \mathfrak{F}_D)$. Designemos por $\mathscr{D}$ el conjunto de todas estas funciones.

El conjunto de estrategias del jugador II (de la naturaleza) Θ queda el anterior.

Si hacemos uso de la decisión $\delta(X)$, y la naturaleza elige θ, nuestras pérdidas constituirán $w(\delta(X), \theta)$. Es una variable aleatoria. Para evitar esta incomodidad, es natural que en calidad de pérdidas para las estrategias $\delta = \delta(\cdot) \in \mathscr{D}$ y $\theta \in \Theta$ se tome el valor de la esperanza matemática

$$W(\delta(\cdot), \theta) = \mathbf{M}_\theta w(\delta(X), \theta) = \int w(\delta(x), \theta)\mathbf{P}_\theta(dx), \qquad (1)$$

que se llama *función de riesgo* (la aparición de la palabra "riesgo" aquí es natural, ya que la aplicación de $\delta(\cdot)$ da un resultado aleatorio). Si se cumple la condición (A_μ) acerca de la existencia de la densidad $f_\theta(x)$ de la distribución $\mathbf{P}_\theta$ con respecto a cierto σ-finita medida μ, entonces la función de riesgo puede escribirse en la forma

$$W(\delta(\cdot), \theta) = \int w(\delta(x), \theta) f_\theta(x)\mu^n(dx).$$

Ahora podemos dar la siguiente

*) En las construcciones de este párrafo podríamos, sin limitar la generalidad, considerar que $n = 1$. Sin embargo, conservaremos el concepto de muestra de volumen n con el fin de dejar válidos los vínculos simples con los resultados de los capítulos precedentes y con las consideraciones posteriores (§§ 6—8).

Una concepción más general de juego estadístico trata de una muestra indefinida $(X_\infty = (x_1, x_2, \ldots)$, en la cual la utilización del elemento x_n va acompañada de las pérdidas $c_n \geq 0$ (véase [63]).

Definición 1. Se llama *juego estadístico* la terna $(\mathscr{D}, \Theta, W)$, donde Θ es el conjunto de estrategias de la naturaleza, $\mathscr{D}$ es el conjunto de todas las aplicaciones medibles del espacio $\mathscr{X}^n$ en el conjunto D, y W ha sido definida en (1). Para caracterizar más completamente el juego estadístico, junto con la terna $(\mathscr{D}, \Theta, W)$ se puede considerar también dado el par $(X, \mathbf{P}_\theta)$, donde $X \in \mathbf{P}_\theta$.

Ejemplo 1. Supongamos que $\theta \in [0, 1]$ determina el contenido de cierto componente químico de la mena preparada para la fusión. Si tomamos la decisión de que la porción de este componente es igual a $\delta \neq \theta$, y de acuerdo con esta decisión se organiza todo el proceso de fusión, entonces, como resultado, la calidad del metal fundido será peor que cuando $\delta = \theta$, y el consumo de energía será más alto. En otros términos, sufriremos las pérdidas $w(\delta, \theta)$ que serán tanto más grandes cuanto más se distinga δ de θ. Supongamos, para abreviar, que $w(\delta, \theta)$ es proporcional al cuadrado de desviación de δ de θ:

$$w(\delta, \theta) = c(\delta - \theta)^2.$$

(Si la función $w(\delta, \theta)$ es suave y si se examina el entorno de la recta $\delta = \theta$, la suposición simplificadora será aquí únicamente la independencia de c respecto a θ). Como resultado obtendremos el juego (D, Θ, w), en el cual $D = [0, 1]$, $\Theta = [0, 1]$,

$$w(\delta, \uparrow) = \sup_\theta w(\delta, \theta) = \begin{cases} c\delta^2 & \text{para } \delta > 1/2, \\ c(1 - \delta)^2 & \text{para } \delta \leqslant 1/2, \end{cases}$$

$$w^* = \inf_\delta w(\delta, \uparrow) = w(1/2, \uparrow) = c/4.$$

Ahora bien, la estrategia $\delta = 1/2$ es minimax y garantiza las pérdidas $\leqslant c/4$. Como $w_* = 0$, este juego no tiene precio. La randomización del juego no mejora la estrategia minimax $\delta = 1/2$ (da $\tilde{w}_* = c/4$). Le dejamos al lector que él mismo se cerciore de que la estrategia bayesiana δ_Q tiene aquí la forma $\delta_Q = \mathbf{M}_Q\theta = \int t\mathbf{Q}(dt)$ (esto resulta de las igualdades $\tilde{w}(\delta, \mathbf{Q}) = c\mathbf{M}_Q(\delta - \theta)^2 = c\mathbf{M}_Q(\theta - \mathbf{M}_Q\theta)^2 + c\mathbf{M}_Q(\delta - \mathbf{M}_Q\theta)^2$) y que la peor distribución $\overline{\mathbf{Q}}$ tendrá la forma $\overline{\mathbf{Q}}(\{0\}) = \overline{\mathbf{Q}}(\{1\}) = 1/2$. Es evidente que la estrategia bayesiana correspondiente es $\delta_{\overline{Q}} = 1/2$.

Supongamos ahora, que la mena es heterogénea y que tenemos la posibilidad de tomar n pruebas de mineral. Estas pruebas se realizan de modo que los resultados de los análisis de laboratorio para el contenido del componente mencionado en las pruebas sean aleatorios y nos den los valores independientes de $(x_1, \ldots, x_n) = X$ respecto a los cuales se sabe que $\mathbf{M}x_i = \theta$, $\mathbf{D}x_i = b/(\theta)$. En este caso, como decisiones $\delta(X)$ servirán todas las estimaciones posibles $\theta^* = \delta(X)$ del parámetro θ según la muestra X. El riesgo de la función de decisión $\delta(X)$ será igual a

$$W(\delta, \theta) = c\mathbf{M}_\theta(\delta(X) - \theta)^2,$$

y llegamos al problema de determinación de la estimación $\theta^* = \delta(X)$ que minimiza en uno u otro sentido este riesgo. Si ponemos, por ejemplo, $\delta_1(X) = \bar{x}$, obtenemos

$$W(\delta_1, \theta) = \frac{cb(\theta)}{n}. \qquad (2)$$

El valor máximo de $b(\theta)$ es igual a $\theta(1-\theta)$ y se alcanza en la distribución x_t concentrada en los puntos 0 y 1.

Como tal posibilidad se puede excluir, entonces

$$b(\theta) < \theta(1-\theta) \leqslant 1/4, \quad W(\delta_1, \theta) < c/4n.$$

Ahora bien, incluso en el caso de $n = 1$ y cuando se utiliza, quizás, no la mejor estrategia, obtenemos un resultado que es mejor que para la estrategia minimax en el juego sin muestra. La relación (2) también indica que el riesgo converge hacia el cero cuando $n \to \infty$. ◁

De la definición dada anteriormente del juego estadístico se deduce que este último posee un conjunto mucho más rico de estrategias $\mathscr{D}$ en comparación con el juego inicial (D, Θ, w).

Al igual que en el § 2, a la par con el juego $(\mathscr{D}, \Theta, W)$, las estrategias del cual llamaremos *puras*, se pueden examinar *juegos randomizados* o *mixtos* $(\tilde{D}, \tilde{\Theta}, \tilde{W})$. Aquí el conjunto $\tilde{\mathscr{D}}$ es el de las aplicaciones de $\pi(X)$: $\mathscr{X}_n \to \tilde{D}$. Estas aplicaciones deben ser tales que los valores

$$\tilde{w}(\pi(X), \theta) = \int_D w(u, \theta)\pi(X, du)$$

sean variables aleatorias; ($\pi(X, A)$ es la probabilidad del conjunto $A \subset D$ en consonancia con la regla de decisión π). Entonces, por definición, ponemos

$$\tilde{W}(\pi(\cdot), \mathbf{Q}) = \int_\Theta \int_{\mathscr{X}^n} \int_D w(u, t)\pi(x, du)\mathbf{P}_t(dx)\mathbf{Q}(dt).$$

La estrategia $\pi(X)$ se llama *regla randomizada de decisión*.

Las relaciones de orden parcial entre las estrategias, las estrategias uniformemente mejores, bayesianas y minimax, y las clases completas para los juegos estadísticos se definen exactamente igual que para los juegos ordinarios (sustituyendo el conjunto D por $\mathscr{D}$, y las funciones w y $\tilde{w}$, por W y $\tilde{W}$).

Las afirmaciones de los teoremas 2.1—2.5 se extienden por completo a los juegos estadísticos, ya que estas afirmaciones de ningún modo están relacionadas con la naturaleza del conjunto D.

2. Clasificación de los juegos estadísticos. Con la naturaleza de los conjuntos D y Θ está vinculada la siguiente clasificación que separa los tipos principales de los juegos estadísticos:

1) Si $\Theta = A$, $D = A$, donde A es un subconjunto "sólido" en R^k (por ejemplo, un paralelepípedo), $w(t, t) = 0$, $w(t, u) > 0$ para $t \neq u$, obtenemos los problemas de la teoría de estimación puntual del parámetro desconocido θ.

2) Si los conjuntos $\Theta = \{\theta_1, \ldots, \theta_r\}$, $D = \{\delta_1, \ldots, \delta_r\}$ son finitos y contienen un número igual de elementos, $w(\delta_i, \theta_i) = 0$, $w(\delta_i, \theta_j) > 0$ para $i \neq j$, obtenemos los problemas de verificación de un número finito de hipótesis simples.

3) Si Θ es una región "sólida" en R^k, $D = \{\delta_1, \delta_2\}$ se compone de dos elementos, $w(\delta_1, \theta) = 0$ para $\theta \in \Theta_1$, $w(\delta_2, \theta) = 0$ para $\theta \in \Theta_2$ ($\Theta_1 \cap \Theta_2$ es un vacío) y $w(\delta_i, \theta) > 0$ en los demás casos, llegamos al problema de verificación de las hipótesis $\{\theta \in \Theta_1\}$ y $\{\theta \in \Theta_2\}$.

Son posibles, desde luego, también otras clases de problemas. Hemos destacado estos tres tipos, puesto que han sido examinados en los capítulos 2 y 3. Además, hemos investigado estos problemas partiendo de posiciones puramente "estadísticas", lo que corresponde a una elección especial de las funciones $w(\delta, \theta)$; en el primer grupo de problemas, las pérdidas se han determinado por la desviación estándar, lo que corresponde a la función de pérdidas $w(\delta, \theta) = (\delta - \theta)^2$; en el segundo grupo, las pérdidas se han determinado por la probabilidad de equivocarse, lo que corresponde a la función

$$w(\delta_i, \theta_j) = \begin{cases} 0, & i = j, \\ 1, & i \neq j. \end{cases}$$

Lo mismo se refiere también al tercer grupo de problemas, en el cual hemos utilizado la función de pérdidas

$$w(\delta_1, \theta) = \begin{cases} 0 & \text{para} \quad \theta \in \Theta_1, \\ 1 & \text{para} \quad \theta \in \Theta_2. \end{cases}$$

$$w(\delta_2, \theta) = \begin{cases} 1 & \text{para} \quad \theta \in \Theta_1. \\ 0 & \text{para} \quad \theta \in \Theta_2. \end{cases}$$

Llamaremos funciones *estadísticas* las funciones de pérdidas que corresponden a un enfoque puramente estadístico de los problemas.

La clasificación citada muestra que no existe ninguna diferencia de principio entre los problemas de la teoría de estimación y la verificación de las hipótesis estadísticas. Todo consiste exclusivamente en la naturaleza de los conjuntos Θ y D y en la forma de las funciones de pérdidas.

Tomando como ejemplo esta clasificación, se puede señalar una peculiaridad más de los juegos estadísticos (en adición a los puntos 1 y 2 dados al principio de este párrafo); esta peculiaridad consiste en que en los juegos estadísticos, el conjunto D ora coincide con Θ ora es un conjunto más pobre que Θ.

3. Dos teoremas fundamentales de la teoría de los juegos estadísticos. Vamos a formular ahora los resultados principales de la teoría de los juegos estadísticos. Ya hemos indicado que las afirmaciones de los teoremas 2.1—2.5 quedan válidas, ya que no están relacionadas con la naturaleza de los juegos. Para obtener dos teoremas fundamentales mencionados en el § 2, introduzcamos ciertas suposiciones. No son, ni mucho menos, las suposiciones más generales (de lo contrario, las enunciaciones y demostraciones se complicarían extraordinariamente), pero son bastante amplias para abarcar el grupo más interesante y sustancial de problemas y, en particular, los examinados en los capítulos 2 y 3.

Condición (A). *Cada uno de los conjuntos* Θ *y* D *es finito o es un conjunto compacto en* R^k.

Como ya hemos señalado, el caso cuando Θ es finito, y $D \subset R^k$, se puede dejar sin examinar. En los demás tres casos vamos a suponer que la función de pérdidas $w(\delta, \theta)$ satisface la condición siguiente.

Condición (B).

1) *Si* $D \subset R^k$, $\Theta \subset R^k$, *la función* $w(\delta, \theta)$ *será continua en* $D \times \Theta$.

2) *Si* $\Theta \subset R^k$ *y* $D = \{\delta_1, \dots, \delta_r\}$ *es finito, cada* r *de las funciones* $w(\delta_i, \theta)$, $i = 1, \dots, r$ *será continua en* Θ.

Si $\Theta = \{\theta_1, \dots, \theta_r\}$ *y* $D = \{\delta_1, \dots, \delta_r\}$ *son finitas, los valores de* $w(\delta_i, \theta_j)$, $i, j = 1, \dots, r$ *pueden ser arbitrarios.*

Además, exigiremos que se cumpla la

Condición (C). *Disponemos de la muestra* $X \in \mathbf{P}_\theta$ *de la distribución* $\mathbf{P}_\theta$, *absolutamente continua para todos* θ *respecto a cierta medida* σ*-finita. Si* $\Theta \subset R^k$, *entonces la densidad* $\frac{d\mathbf{P}_\theta}{d\mu}(x) = f_\theta(x)$ *es continua en* L_1 $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mu)$ *respecto a* θ, *o sea, para* $\theta_m \to \theta$,

$$\int |f_{\theta_m}(x) - f_\theta(x)|\mu(dx) \to 0. \tag{3}$$

No es difícil comprobar que la continuidad ordinaria $f_\theta(x)$ respecto a θ, para $[\mu]$ c.t. x, contribuye a la continuidad (3).

Teorema 1. *Si se cumplen las condiciones* (A), (B), (C), *el juego promediado* $(\mathscr{D}, \bar{\Theta}, \bar{W})$ *tiene precio y estrategias minimax* $\bar{\pi}(X)$ y $\bar{\mathbf{Q}}$:

$$\bar{W}(\bar{\pi}(\cdot), \uparrow) = \inf_{\pi} \bar{W}(\pi(\cdot), \uparrow), \quad \bar{W}(\downarrow, \bar{\mathbf{Q}}) = \sup_{\mathbf{Q}} \bar{W}(\downarrow, \mathbf{Q}).$$

De los teoremas 2.4 y 2.5 del párrafo precedente sabemos que $\bar{\mathbf{Q}}$ es la peor distribución,

$$\bar{W}(\pi_{\bar{\mathbf{Q}}}(\cdot), \bar{\mathbf{Q}}) = \sup_{\mathbf{Q}} \bar{W}(\pi_{\mathbf{Q}}(\cdot), \mathbf{Q}) = \sup_{\mathbf{Q}} \bar{W}(\downarrow, \mathbf{Q}),$$

y $\bar{\pi}(X) = \pi_{\bar{\mathbf{Q}}}(X)$ es la estrategia bayesiana correspondiente a $\bar{\mathbf{Q}}$.

Sabemos también (véase el teorema 2.5) que para que la estrategia $\bar{\pi}(X)$ sea minimax, es necesario y suficiente que la misma sea bayesiana:

$\bar{\pi}(X) = \pi_Q(X)$ para cierta distribución a priori **Q**, y

$$\tilde{W}(\bar{\pi}(\cdot), \bar{\theta}) = c = \text{const c.d } [\mathbf{Q}],$$
$$\tilde{W}(\bar{\pi}(\cdot), \tilde{\theta}) \leqslant c.$$

Este último criterio del carácter minimax ya fue utilizado reiteradas veces en diferentes situaciones particulares (véanse los §§ 2.11, 3.1, 3.5 y 3.9).

Teorema 2. *Al cumplirse las condiciones* (A), (B), (C), *la clase de todas las estrategias bayesianas será completa.*

En el Suplemento VIII aducimos las demostraciones de los teoremas 1 y 2 en su forma más general, cuando D y Θ son espacios métricos compactos arbitrarios (condición (A)); la función $w(\delta, \theta)$: $D \times \Theta \to R$ es continua respecto a δ y θ en las métricas respectivas (condición (B)); la distribución $\mathbf{P}_\theta$ es continua respecto a θ según la variación (condición (C)).

Las demostraciones de los teoremas 1 y 2 en caso de ciertas suposiciones adicionales, se pueden deducir de [90]. Sin embargo, las demostraciones para el caso de D y Θ finitos se pueden deducir de [7] y [93]. En estas mismas monografías es posible hallar una exposición relativamente completa de los elementos de la teoría general de los juegos estadísticos (y, en particular, la investigación para algunos casos de construcción de la clase completa mínima; véase [93]).

Los teorema 1 y 2 muestran cuán importante es el problema de descripción de la clase de todas las reglas bayesianas de decisión. El siguiente párrafo está dedicado a este problema.

§ 4. Principio bayesiano. Clase completa de funciones de decisión

Hemos visto que por su construcción el juego estadístico es un objeto más complejo que el juego inicial (D, Θ, w). Para este juego, sobre todo si se trata de los conjuntos simples D y Θ (por ejemplo, finitos), la determinación de las estrategias bayesianas y minimax puede ser una tarea relativamente sencilla. Al mismo tiempo, incluso el conjunto D de los juegos estadísticos elementales es de naturaleza muy compleja, y esto puede dificultar considerablemente el estudio de dichos juegos, siempre que los mismos se consideren como juegos ordinarios.

Ejemplo 1. Supongamos que los conjuntos $D = \{\delta_1, \delta_2\}$, $\Theta = \{\theta_1, \theta_2\}$ son bipuntuales, $w(\delta_i, \theta_j) = w_{ij}$, $w_{ii} = 0$, $i, j = 1, 2$. Sea $\mathbf{Q} = (q, 1 - q)$ la distribución a priori en Θ. Entonces,

$$\tilde{w}(\delta_i, \mathbf{Q}) = q w_{i1} + (1 - q) w_{i2}.$$

Por consiguiente, la estrategia bayesiana π_Q tiene la forma

$$\pi_Q(\delta_2) = \begin{cases} 0, & \text{si} \quad \tilde{w}(\delta_1, \mathbf{Q}) < \tilde{w}(\delta_2, \mathbf{Q}) \quad (q w_{21} > (1 - q) w_{12}), \\ 1, & \text{si} \quad \tilde{w}(\delta_2, \mathbf{Q}) \quad \tilde{w}(\delta_1, \mathbf{Q}) \quad (q w_{21} < (1 - q) w_{12}) \end{cases} \qquad (1)$$

($\pi_Q(\delta_i)$ es la probabilidad de que se acepte δ_i).

Si

$$\tilde{w}(\delta_1, \mathbf{Q}) = \tilde{w}(\delta_2, \mathbf{Q}) \tag{2}$$

o bien, que es lo mismo, si $q = q = w_{12} + (w_{12} + w_{21})$, entonces, en calidad de π_Q se puede tomar cualquier distribución de π en el conjunto (δ_1, δ_2). De un modo exactamente igual siempre se puede hallar una distribución de $\pi = (p, 1 - p)$ tal, que

$$\tilde{w}(\pi, \theta_1) = \tilde{w}(\pi, \theta_2), \text{ o bien } pw_{12} = (1 - p)w_{21}.$$

La solución de esta ecuación $\bar{p} = w_{21}/(w_{21} + w_{12})$ responde, evidentemente, a la estrategia bayesiana igualadora $\pi_{\bar{Q}}$, $\bar{\mathbf{Q}} = (\bar{q}, 1 - \bar{q})$, la cual, en virtud de los teoremas 2.4 y 2.5, será minimax. La distribución $\bar{\mathbf{Q}}$ será la peor.

Vemos que la "resolución" de este juego se lleva a cabo bastante simplemente. No obstante, si se pasa al juego estadístico, incluso en el caso elemental de $w_{12} = w_{21} = 1$, obtendremos el problema de los criterios bayesianos y minimax para cuya investigación hemos necesitado dos párrafos: 3.1 y 3.2.

Un hecho magnífico, al cual dedicamos el presente párrafo, consiste en que el problema de determinación de las estrategias bayesianas (y, por lo tanto, de la clase completa y de las estrategias minimax) para los juegos estadísticos puede ser reducido, en cierto sentido, al mismo problema para los juegos iniciales (D, Θ, w). Esta reducción se basa en la afirmación siguiente, la cual llamaremos *principio bayesiano*: Sea, como antes,

$$f_\theta(X) = \prod_{i=1}^{n} f_\theta(x_i)$$

la función de verosimilitud de la muestra X y sea ella misma la densidad de X en $\mathscr{X}^n$ respecto a μ^n. Supongamos, además, que la distribución a priori $\mathbf{Q}$ en $(\Theta, \mathfrak{F}_\Theta)$ tiene una densidad $q(t)$ respecto a cierta medida λ (es evidente que esto no es una limitación). Entonces, de acuerdo con el § 2.11, la función $f(x, t) = q(t)f_t(x)$ será la densidad de la distribución compatible de (X, θ) en $\mathscr{X}^n \times \Theta$. Esto quiere decir que la función

$$q(t/x) = \frac{q(t)f_t(x)}{f(x)}, \tag{3}$$
$$f(x) = \int q(t)f_t(x)\lambda(dt),$$

define la densidad condicional de la distribución de θ a condición de que $X = x$. Esta densidad corresponde a la distribución a posteriori $\mathbf{Q}_x$ de la variable aleatoria θ a condición de que $X = x$. La relación (3) se denomina *fórmula de Bayes* (véanse los §§ 2.10 y 2.11).

Teorema 1 (principio bayesiano). *Supongamos que se cumple la condición* (A_μ), *que la distribución a priori en* Θ *tiene una densidad de* $q(t)$,

y *que* Q_x, significa la distribución a posteriori de densidad (3), *la cual corresponde a la distribución a priori* Q. *Supongamos, además, que el juego inicial* (D, Θ, w) *para cualquier distribución a priori* Q, *tiene la estrategia bayesiana* π_Q. *Entonces, el juego estadístico* $(\mathcal{D}, \Theta, W)$ *tiene una estrategia bayesiana* $\pi_Q(X)$ *correspondiente a la distribución* Q, *la cual coincide con* π_{Q_x}, *o sea, con la estrategia bayesiana del juego inicial, correspondiente a la distribución a posteriori* Q_x.

La afirmación de este teorema se puede expresar por una sola igualdad

$$\pi_Q(X) = \pi_{Q_x}.$$

Esta reduce el problema planteado, al problema de determinación de la distribución a posteriori Q_x y al problema de determinación de las estrategias bayesianas para el juego inicial.

El teorema 1 es muy importante para comprender el mecanismo de influencia de la información obtenida de la muestra, sobre la elección de la estrategia óptima. La información a priori, representada por la distribución Q en Θ, varía continuamente bajo la influencia de los datos experimentales. La estrategia óptima será la que tendrá en cuenta estas variaciones, del modo siguiente: es necesario tomar la estrategia óptima en el juego inicial, pero que ya no corresponde a Q, sino a Q_x.

Demostración del teorema 1. Tenemos

$$\tilde{W}(\pi(\cdot), Q) = \int_{\Theta}\int_{\mathcal{X}^n} \tilde{w}(\pi(x), t) f_t(x)\mu^n(dx) q(t)\lambda(dt) =$$

$$= \int_{\mathcal{X}^n} f(x)\mu^n(dx) \int_{\Theta} \tilde{w}(\pi(x), t) q(t/x)\lambda(dt). \quad (4)$$

Aquí hemos utilizado (3). El cambio del orden de integración es justo en virtud del carácter no negativo de la función subintegral. La segunda integral en el segundo miembro (4) no es otra cosa sino $\tilde{w}(\lambda(x), Q_x)$. Pero para cualquier x,

$$\tilde{w}(\pi(x), Q_x) \geqslant \tilde{w}(\pi_{Q_x}, Q_x) = \int_{\Theta} \tilde{w}(\pi_{Q_x}, t) q(t+x)\lambda(dt).$$

Sustituyendo esta desigualdad en (4) y volviendo al orden inicial de integración, obtenemos

$$\tilde{W}(\pi(\cdot), Q) \geqslant \int_{\mathcal{X}^n} f(x)\mu^n(dx) \int_{\Theta} \tilde{w}(\pi_{Q_x}, t) q(t+x)\lambda(dt) = \tilde{W}(\pi_Q, Q).$$

En vista de que aquí $\pi(x)$ es arbitraria, esto quiere decir que

$$\pi_Q(x) = \pi_{Q_x}. \lhd$$

Observación 1. Con fines de precisión, en las consideraciones citadas debemos especificar la mensurabilidad de la función $\tilde{w}(\pi_{Q_x}, t)$ respecto a

$\mathfrak{B}^n \times \mathfrak{F}_\Theta$. Omitimos estas restricciones, ya que éstas tienen un carácter puramente técnico y, al cumplirse las condiciones (A), (B) y (C) del § 3, son completamente innecesarias. El lector puede comprobar personalmente esta última afirmación, utilizando el hecho de que para D y Θ discretos, tal mensurabilidad se establece de un modo evidente, así como el hecho de que el juego arbitrario, al cumplirse las condiciones (A) y (B), puede ser "aproximado" al juego discreto tan exactamente como se quiera.

Volviendo al ejemplo 1, ahora podemos, en virtud del teorema 1, señalar inmediatamente el tipo de estrategias bayesianas para el juego estadístico respectivo. Precisamente de (1) obtenemos

$$\pi_{Q_x}(\delta_2) = \begin{cases} 0, & \text{si} \quad q_x = \dfrac{qf_{\theta_1}(X)}{qf_{\theta_1}(X) + (1-q)f_{\theta_2}(X)} > \dfrac{w_{12}}{w_{12}+w_{21}}, \\ 1, & \text{si} \quad q_x < \dfrac{w_{12}}{w_{12}+w_{21}}. \end{cases} \tag{5}$$

Si

$$q_x = \frac{w_{12}}{w_{12}+w_{21}}, \tag{6}$$

entonces, en calidad de π_{Q_x} se puede tomar cualquier distribución en (δ_1, δ_2). La desigualdad (5) se puede escribir de la forma siguiente:

$$\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}, \quad a = \frac{w_{12}}{w_{12}+w_{21}}. \tag{7}$$

Este es el criterio de relación de verosimilitud que ya conocemos.

Seguidamente,

$$\bar{W}(\pi_Q, \theta_j) = w_{1j}\mathbf{M}_{\theta_j}\pi_{Q_x}(\delta_1) + w_{2j}\mathbf{M}_{\theta_j}\pi_{Q_x}(\delta_2), \quad j = 1, 2.$$

Supongamos, para abreviar, que la igualdad (6) tiene lugar con $\mathbf{P}_{\theta_j}$ probabilidad de 0, así que la estrategia bayesiana con $\mathbf{P}_{\theta_j}$ probabilidad de 1 será pura, $j = 1, 2$. Entonces,

$$\mathbf{M}_{\theta_j}\pi_{Q_x}(\delta_1) = \mathbf{P}_{\theta_j}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}\right).$$

$$\bar{W}(\pi_Q, \theta_1) = w_{21}\mathbf{P}_{\theta_1}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} < \frac{a(1-q)}{q(1-a)}\right).$$

$$\bar{W}(\pi_Q, \theta_2) = w_{12}\mathbf{P}_{\theta_2}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}\right).$$

De aquí ya no es difícil hallar el valor de $\bar{q}$ correspondiente a la peor distribución $\bar{Q}$ para el cual $\pi_{\bar{Q}_x}$ será la estrategia igualadora, o sea, la estrategia con la que

$$\bar{W}(\pi_{\bar{Q}}, \theta_1) = \bar{W}(\pi_{\bar{Q}}, \theta_2).$$

Según los teoremas 2.4 y 2.5, esta estrategia será minimax. Le dejamos al lector que él mismo extienda el procedimiento descrito de determinación de la estrategia minimax, al caso general cuando $\mathbf{P}_{\theta_1}$- o $\mathbf{P}_{\theta_2}$-distribuciones $f_{\theta_1}(X)/f_{\theta_2}(X)$ contienen la componente discreta.

Valiéndonos del teorema 1 podemos, de un modo análogo, obtener la generalización de los resultados de los §§ 3.1 y 3.2 para el caso de D y Θ finitos arbitrarios y de una función arbitraria de pérdidas $w(\delta_i, \theta_j) = w_{ij}$, la cual en este caso también puede llamarse matriz de pérdidas $\|w(\delta_i, \theta_j)\|$. (En los párrafos §§ 3.1 y 3.2 hemos examinado el caso particular de $w_{ij} = 1$ cuando $i \neq j$). Para w_{ij} arbitrarias, la regla bayesiana de decisión tendrá la forma siguiente. Sea $\mathbf{Q} = (q(\theta_1), \ldots, q(\theta_r))$, $\mathbf{Q}_x = (q_X(\theta_1), \ldots, q_X(\theta_r))$,

$$q_X(\theta_j) = \frac{q(\theta_j) f_{\theta_j}(X)}{\sum_i q(\theta_i) f_{\theta_i}(X)}.$$

Entonces, $\tilde{w}(\delta_i, \mathbf{Q}_X) = \sum_{j=1}^{r} w_{ij} q_X(\theta_j)$ y, por lo tanto,

$\pi_{Q_X}(\delta_k) = 1$, si $\tilde{w}(\delta_k, \mathbf{Q}_X) \leqslant \tilde{w}(\delta_i, \mathbf{Q}_X)$ para todos i, o bien, que es lo mismo, si

$$\sum_{j=1}^{r} w_{kj} f_{\theta_j}(X) q(\theta_j) \leqslant \sum_{j=1}^{r} w_{ij} f_{\theta_j}(X) q(\theta_j).$$

Si existen varios valores de k dotados de esta propiedad (designémoslos por $k_1, \ldots, k_s$), entonces, cualquier distribución en $\delta_{k_1}, \ldots, \delta_{k_s}$ también será una estrategia bayesiana π_{Q_X}.

La determinación de la estrategia minimax se lleva a cabo del modo siguiente. Supongamos, también para abreviar, que $\mathbf{P}_{\theta_j}$-distribuciones $\tilde{w}(\delta_i, \mathbf{Q}_X)$ no tienen componentes discretas. Entonces,

$$\tilde{W}(\pi_{Q\cdot}, \theta_j) = \sum_{i \neq j} w_{ij} \mathbf{P}_{\theta_j}(\tilde{w}(\delta_i, \mathbf{Q}_X) < \min_{l \neq i} \tilde{w}(\delta_l, \mathbf{Q}_X)).$$

En virtud del teorema 3.1 existe $\bar{\mathbf{Q}} = (\bar{q}(\theta_1), \ldots, \bar{q}(\theta_r))$ con la que la estrategia $\pi_{\bar{Q}_X}$ igualará los valores de $\tilde{W}(\pi_{\bar{Q}_X}, \theta_j)$ para todos los valores de j. Esta estrategia será precisamente minimax.

De las consideraciones citadas y del teorema 3.2 también es fácil obtener el tipo de clase completa de estrategias del juego estadístico $(\mathscr{D}, \Theta, \tilde{W})$ en el caso de D y Θ finitos.

Examinemos las estrategias π_{Q_X} que son la distribución aleatoria de tales $\delta_{k_1}, \ldots, \delta_{k_s}$ para los cuales

$$\min_{l} \left(\sum_{j=1}^{r} (w_{k_l, j} - w_{lj}) f_{\theta_j}(X) q(\theta_j) \right) = 0.$$

La clase de tales estrategias (bayesianas), que se obtienen si $q(\theta_1, \ldots, q(\theta_r)$, recorrerán todos los valores posibles y serán una clase completa. Hemos visto que en el caso de $r = 2$ esta clase resulta muy simple y estrecha (véase (7)): consta de las funciones de decisión $\pi(X) = (\pi(X, \delta_1), \pi(X, \delta_2))$, donde $\pi(X, \delta_i)$ son las probabilidades de que se tome la decisión δ_i,

$$\pi(X, \delta_1) = \begin{cases} 1, & \text{si} \quad R(X) > c, \\ p \in [0, 1], & \text{si} \quad R(X) = c, \\ 0 & \text{si} \quad R(X) < c, \end{cases}$$

$$R(X) = \frac{f_{\theta_1}(X)}{f_{\theta_2}(X)}, \quad 0 \leqslant c \leqslant \infty. \tag{8}$$

En los juegos continuos con conjuntos D y Θ para algunas funciones de pérdidas concretas importantes también es posible hallar la forma explícita de las decisiones bayesianas. Supongamos, por ejemplo, que D y Θ son las regiones de R^k, y que la función de pérdidas es cuadrática:

$$w(\delta, \theta) = c|\delta - \theta|^2 = c \sum_{i=1}^{k} |\delta_i - \theta_i|^2, \tag{9}$$

donde δ_i, θ_i son las coordenadas δ y θ. Entonces,

$$\tilde{w}(\delta, \mathbf{Q}) = c \int |\delta - t|^2 \mathbf{Q}(dt) = c\mathbf{M}_Q|\delta - \theta|^2.$$

Sabemos que el mínimo de esta expresión se alcanza para $\delta = \mathbf{M}_Q\theta = \int t\mathbf{Q}(dt)$. Esto es, evidentemente, la estrategia bayesiana $\delta_Q = \mathbf{M}_Q\theta$. De aquí y del principio bayesiano resulta que la estrategia bayesiana $\delta_Q(X) = \theta_Q^*$ en el juego estadístico tendrá la forma siguiente:

$$\theta_Q^* = \delta_{Q_x} = \int_{R^k} t\mathbf{Q}_x(dt) = \int_{R^k} tq(t/X)\lambda(dt). \tag{10}$$

Este resultado ya fue obtenido en el capítulo 2.

El riesgo de la estrategia θ_Q^* es igual a $W(\theta_Q^*, \theta) = c\mathbf{M}_\theta|\theta_Q^* - \theta|^2$. La distribución a priori $\bar{\mathbf{Q}}$, para la cual $\mathbf{M}_\theta|\theta_{\bar{Q}}^* - \theta|^2 = \text{const}$, nos ofrecerá la estimación minimax $\bar{\theta}^* = \delta_{\bar{Q}}(X)$. Ejemplos de construcción de estimaciones minimax en esta vía se dan en el § 2.11.

La clase de estimaciones (10), donde $\mathbf{Q}$ recorre los valores en la clase de todas las distribuciones en Θ, es una clase completa.

Examinemos ahora otro caso particular de la función de pérdidas

$$w(\delta, \theta) = c|\delta - \theta| \tag{11}$$

y supongamos que $\Theta = R$, $D = R$. Entonces,

$$\tilde{w}(\delta, \mathbf{Q}) = c\mathbf{M}_Q|\delta - \theta| = c\int |\delta - t|\mathbf{Q}(dt) =$$

$$= c\int_{-\infty}^{\delta} (\delta - t)\mathbf{Q}(dt) + c\int_{\delta}^{\infty} (t - \delta)\mathbf{Q}(dt).$$

Utilizando la integración por partes y designando $F(t) = \mathbf{Q}((-\infty, t))$, hallamos

$$\bar{w}(\delta, \mathbf{Q}) = c \int_{-\infty}^{\delta} (\delta - t)dF(t) - c \int_{\delta}^{\infty} (t - \delta)d(1 - F(t)) =$$

$$= c\left[\int_{-\infty}^{\delta} F(t)dt + \int_{\delta}^{\infty} (1 - F(t))dt\right].$$

La derivada de esta expresión respecto a δ existe c.d. y es igual a $c[2F(\delta) - 1]$. Esta función crece monótonamente y cambia de signo en el punto $\bar{\delta}$, igual a la mediana de la distribución F: $F(\bar{\delta} - 0) \leqslant 1/2$, $F(\bar{\delta} + 0) \geqslant 1/2$. De aquí se deduce que $w(\delta, \mathbf{Q})$ será convexa en cuanto a δ y en el punto $\bar{\delta}$ tendrá el mínimo valor.

En virtud del principio bayesiano esto quiere decir que la *mediana de la distribución a posteriori* $\mathbf{Q}_X$ *será la estimación bayesiana* $\theta^*_{\mathbf{Q}} = \delta_{\mathbf{Q}}(X)$ *para la distribución a priori* $\mathbf{Q}$ *y la función de pérdidas* (11). Al igual que en el caso (9), esto da la posibilidad de hallar la función de decisión minimax y la clase completa.

De un modo análogo se puede examinar el caso

$$w(\delta, \theta) = c|\delta - \theta|^{\alpha}, \quad \alpha > 0.$$

En conclusión de este párrafo nótese que la función cuadrática de pérdidas (9) en caso de $c = 1$ para los conjuntos continuales D y Θ y la función de pérdidas

$$w(\delta_i, \theta_j) = \begin{cases} 0, & i = j, \\ 1, & i \neq j \end{cases} \tag{12}$$

para D y Θ finitos desempeñan un papel especial en la teoría de los juegos estadísticos. En este caso las funciones de riesgo se convierten en la suma de la varianza y el cuadrado del desplazamiento de la estimación para D y Θ continuales, así como en la probabilidad de equivocarse para D y Θ finitos, respectivamente. Estas características, que son naturales de por sí, nos servían de base para elegir las reglas óptimas en los capítulos 2, 3 y 4. Si un problema estadístico no contiene indicaciones directas concernientes a la forma de la función $w(\delta, \theta)$, entonces con más frecuencia en calidad de $w(\delta, \theta)$ se eligen precisamente estas dos funciones: (9) ó (12). Hemos decidido llamarlas *funciones estadísticas de pérdidas.*

§ 5. Suficiencia, carácter no desplazado e invariación

Los principios de suficiencia, de carácter no desplazado y de invariación sirven para reducir la clase de reglas de decisión. Los mismos consisten en utilizar en calidad de funciones de decisión sólo las reglas de decisión

suficientes, no desplazadas e invariantes, respectivamente. La utilización de uno de estos principios, de dos de ellos o de los tres a la vez (si esto es posible) permite, en una serie de casos, reducir hasta tal punto la clase de estrategias sometidas a examen, que su intersección con la clase completa resulta integrada por una sola función de decisión. Esto quiere decir que en la subclase separada existe una estrategia uniformemente mejor (compárese con el punto 1 del § 2) y esto resuelve el problema de elección de la decisión.

Los tres principios son bastante naturales y ya han sido analizados en distintos casos concretos de los capítulos 2 y 3.

El más irrefutable de ellos es el principio de suficiencia, que a menudo no es otra cosa sino el método de descripción de una clase completa.

1. Suficiencia. Supongamos que se cumple la condición (A_μ) y que existe la estadística suficiente S, o sea (véase el § 2.12),

$$f_\theta(X) = \psi(\theta, S)\cdot h(X).$$

Supongamos, además, que la distribución a priori Q tiene una densidad $q(t)$ respecto a cierta medida λ. Entonces, en virtud del principio bayesiano, la estrategia bayesiana será totalmente determinada por la densidad a posteriori

$$q(t/X) = \frac{q(t)f_t(X)}{\int q(u)f_u(X)\lambda(du)} = \frac{q(t)\psi(t, S)}{\int q(u)\psi(u, S)\lambda(du)}, \qquad (1)$$

que depende exclusivamente de S. Como cualquier distribución Q tiene densidad respecto a una medida λ seleccionada respectivamente (se puede poner, por ejemplo, $\lambda = Q$, $q(t) \equiv 1$), lo dicho significa que todas las reglas bayesianas de decisión $\pi_Q(X)$ serán sólo funciones de S:

$$\pi_Q(X) = p_Q(S).$$

Con otras palabras, cualquier estrategia bayesiana $\pi_Q(X)$ no depende de X al ser fija S.

Ahora supongamos que se cumplen las condiciones (A), (B) y (C) del § 3. Entonces, la afirmación enunciada también atañerá a las estrategias minimax. Esto también significará que todas las reglas de decisión construidas tan sólo como funciones de S (o sea, todas las aplicaciones medibles de Φ / $\bar{D}$, donde Φ es el espacio en que se hallan los valores de S), forman la clase completa $\mathscr{D}_\Phi$. Esto se deduce del hecho de que $\mathscr{D}_\Phi$ contiene todas las estrategias bayesianas que forman, como sabemos, la clase completa. Evidentemente, la clase $\mathscr{D}_\Phi$ será la mínima para la estadística suficiente mínima S.

Está claro que la clase completa mínima no comprende todas las funciones de S (con valores en $\bar{D}$), sino tan sólo una parte reducida de las mismas. Eso lo confirma la fórmula (1), de la cual resulta, por ejemplo,

que para los conjuntos bipuntuales D y Θ (véase (4.8)), la clase completa está formada por funciones $\pi(X)$ cuya probabilidad $\pi(X, \delta_1)$ de toma de decisión δ_1 tiene forma de indicador del conjunto $\{R(X) > c\}$, donde $R(X) = \psi(\theta_1, S)/\psi(\theta_2, S)$ (véase, para precisar, (4.8)).

Si $D \subset R^k$, $\Theta \subset R^k$ y la función de pérdidas $w(\delta, \theta)$ tiene la forma $w(\delta, \theta) = w(\delta - \theta)$, donde $w(u)$ es una función convexa en R^k, al principio de suficiencia se le puede conferir una forma muy constructiva que permite caracterizar eficientemente la clase completa, o sea, tiene lugar la siguiente generalización del teorema 2.14.1.

Teorema 1 (Blackwell). *Para cualquier función de decisión (estimación) $\theta^* = \delta(X)$ existe la estimación*

$$\theta_S^* = \mathbf{M}_\theta(\theta^*/S)$$

(θ_S^* no depende de θ, ya que S es una estadística suficiente) la cual no es peor que θ^*, o sea, para todos $\theta \in \Theta$,

$$\mathbf{M}_\theta w(\theta_S^* - \theta) \leqslant \mathbf{M}_\theta w(\theta^* - \theta).$$

Demostración. Tiene lugar la siguiente desigualdad de Jensen (véase el § 2.9): si g es una función convexa en R^k; ξ, una variable aleatoria con valores en R^k; y $\mathfrak{F}$, cualquier σ-subálgebra de la σ-álgebra principal, entonces

$$\mathbf{M}(g(\xi)/\mathfrak{F}) \geqslant g(\mathbf{M}(\xi/\mathfrak{F})).$$

Conforme a esta desigualdad,

$$\mathbf{M}_\theta w(\theta^* - \theta) = \mathbf{M}_\theta\{\mathbf{M}_\theta(w(\theta^* - \theta)/S)\} \geqslant$$
$$\geqslant \mathbf{M}_\theta w(\mathbf{M}_\theta(\theta^* - \theta/s)) = \mathbf{M}_\theta w(\theta_S^* - \theta). \lhd$$

Si la estadística suficiente S es completa, el teorema 1, junto con el principio de no desplazamiento, permite determinar unívocamente la mejor estimación. En efecto, examinemos la clase K_0 de todas las estimaciones no desplazadas $\theta^* = \delta(X)$:

$$\mathbf{M}_\theta\theta^* = \theta \quad \text{para} \quad \theta^* \in K_0.$$

Entonces, siguiendo exactamente los razonamientos del § 2.14 (teorema 3), nos convencemos de que $\theta_S^* = \mathbf{M}_\theta(\theta^*/S)$ coinciden para todas $\theta^* \in K_0$ y, por consiguiente, la intersección de K_0 y de la clase completa se compone de una sola estimación $\varphi(S)$, la cual es natural llamarla *eficiente*.

De lo dicho se deduce que las *estimaciones eficientes, si existen, serán las mismas para una función convexa arbitraria de pérdidas* $w(\delta - \theta)$. Esto permite utilizar, para cualquiera de estas funciones, todas las afirmaciones de los teoremas respectivos del capítulo 2, obtenidos para $w(u) = u^2$.

Los razonamientos citados ilustran la aplicación compatible de los principios de suficiencia y de carácter no desplazado.

2. Carácter no desplazado. Acabamos de ver qué papel puede desempeñar el principio de carácter no desplazado en la teoría de las estimaciones. En el § 3.6 hemos establecido que un efecto análogo (existencia de criterios no desplazados uniformemente más potentes) puede obtenerse al utilizar los criterios no desplazados en la teoría de verificación de las hipótesis estadísticas.

En el caso general, el carácter no desplazado se define del modo siguiente. Admitamos que el problema de una decisión estadística consiste en "determinar" el valor desconocido de θ y que, por consiguiente, los conjuntos D y Θ coinciden. La función de pérdidas $w(\delta, \theta)$ puede ser arbitraria.

Definición 1. La función de decisión $\delta(X)$ se llama no desplazada si

$$\mathbf{M}_\theta w(\delta(X), \theta) \leqslant \mathbf{M}_\theta w(\delta(X), \theta')$$

para todos θ, $\theta' \neq \theta$.

Con otras palabras, para $v = \theta$ se alcanza $\min\limits_v \mathbf{M}_\theta w(\delta(X), v)$. Esto significa que $\delta(X)$, por término medio, se encuentra más cerca de θ desconocido que de cualquier otro punto.

Es fácil notar que la definición de las estimaciones no desplazadas que hemos dado anteriormente es un caso particular de esta afirmación.

Si se verifican dos hipótesis compuestas, $H_1 = \{\theta \in \Theta_1\}$ y $H_2 = \{\theta \in \Theta_2\}$, el conjunto $D = \{\delta_1, \delta_2\}$ puede distinguirse considerablemente de Θ. En este caso, la definición del carácter no desplazado será formalmente algo diferente, aunque su sentido queda invariable, o sea, la definición 1 se puede modificar de tal modo (véase [57]) que la misma pase a la definición siguiente.

Definición 1A. La función de decisión $\delta(X)$ se llama *no desplazada* si

$$\mathbf{M}_\theta w(\delta(X), \theta) \leqslant \mathbf{M}_\theta w(\delta(X), \theta')$$

para todos $\theta \in \Theta_1$, $\theta' \in \Theta_2$ o bien $\theta \in \Theta_2$, $\theta' \in \Theta_1$.

Supongamos, para abreviar, que $w(\delta_1, \theta) = w_1 = \text{const}$ para $\theta \in \Theta_2$; $w(\delta_2; \theta) = w_2 = \text{const}$ para $\theta \in \Theta_1$; $\delta_1 = 0$, $\delta_2 = 1$, y que $\delta(X)$ significa la probabilidad (1 ó 0) de que se acepte H_2. Entonces,

$$\mathbf{M}_\theta w(\delta(X), \theta) = \begin{cases} w_2 \mathbf{P}_\theta(\delta(X) = 1) & \text{para} \quad \theta \in \Theta_1, \\ w_1 \mathbf{P}_\theta(\delta(X) = 0) & \text{para} \quad \theta \in \Theta_2, \end{cases}$$

$$\mathbf{M}_\theta w(\delta(X), \theta') = \begin{cases} w_1 \mathbf{P}_\theta(\delta(X) = 0) & \text{para} \quad \theta \in \Theta_1, \ \theta' \in \theta_2, \\ w_2 \mathbf{P}_\theta(\delta(X) = 1) & \text{para} \quad \theta \in \Theta_2, \ \theta' \in \Theta_1 \end{cases}$$

y las desigualdades en la definición 1A quieren decir que

$$w_2 \mathbf{P}_{\theta_1}(\delta(X) = 1) \leqslant w_1 \mathbf{P}_{\theta_1}(\delta(X) = 0) \quad \text{para} \quad \theta_1 \in \Theta_1,$$
$$w_1 \mathbf{P}_{\theta_2}(\delta(X) = 0) \leqslant w_2 \mathbf{P}_{\theta_2}(\delta(X) = 1) \quad \text{para} \quad \theta_2 \in \Theta_2,$$

o bien, que es lo mismo,

$$\mathbf{P}_{\theta_1}(\delta(X) = 1) \leqslant \frac{w_1}{w_1 + w_2}, \quad \mathbf{P}_{\theta_2}(\delta(X) = 1) \geqslant \frac{w_1}{w_1 + w_2}.$$

De aquí se deduce que

$$\sup_{\theta\in\Theta_1} \mathbf{M}_\theta\delta(X) \leqslant \inf_{\theta\in\Theta} \mathbf{M}_\theta\delta(X)$$

y que, por consiguiente, el criterio δ no será desplazado desde el punto de vista de las definiciones del § 3.6. Al contrario, si es válida la última desigualdad, el criterio δ no será desplazado desde el punto de vista de la definición 1A al elegir adecuadamente la función de pérdidas $w(\delta, \theta)$, por ejemplo, para $w_1/(w_1 + w_2) = \sup_{\theta\in\Theta_1} \mathbf{M}_\theta\delta(X)$.

Los ejemplos adicionales de utilización del principio no desplazado (además de los resultados obtenidos en el § 3.6) se pueden hallar en [57].

3. Invariación. Hemos visto que la intersección de la clase completa, engendrada por las decisiones "suficientes", con la clase de decisiones no desplazadas puede constar de una sola estrategia. La clase de reglas de decisión invariantes es otra clase natural de estrategias, en la que puede resultar la única decisión inmejorable (compárese con los §§ 2.18, 2.19 y 3.7).

La definición del problema invariante de decisión estadística está relacionada con los grupos de transformaciones en los tres espacios que participan en la definición del juego estadístico: en los espacios D y Θ y en el espacio muestral $\mathscr{X}^n$. La definición se basa en las transformaciones biunívocas medibles g del espacio $\mathscr{X}^n$ en sí, que forman cierto grupo G con la operación de grupo definida como una composición: si $g_1 \in G$ y $g_2 \in G$, entonces g_2g_1 se define como una transformación $x \to g_2(g_1x)$ que otra vez debe pertenecer a G. Designemos por e la transformación idéntica. Sin embargo, la transformación g^{-1} inversa a g se define como una transformación para la cual $g^{-1}g = e$. La mensurabilidad de $g \in G$ significa que gX, junto con X, será una variable aleatoria en $\mathscr{X}^n$.

Con el grupo introducido G está estrechamente relacionado el concepto de invariación de la familia $\mathbf{P}_\theta$ que hemos definido en los §§ 2.19 y 3.7. Este concepto significa que para $g \in G$ y $\theta \in \Theta$ habrá un elemento $\theta_g \in \Theta$ tal, que

$$\mathbf{P}_\theta(gX \in A) = \mathbf{P}_{\theta_g}(X \in A). \tag{2}$$

Las transformaciones $\bar{g}$ del espacio Θ en sí, definidas por la igualdad $\bar{g}\theta = \theta_g$, al cumplirse la condición (A_0) forman el grupo $\bar{G}$ (véase el § 2.19).

En términos de las esperanzas matemáticas, la condición (2) significa que para cualquier función integrable φ,

$$\mathbf{M}_\theta\varphi(gX) = \mathbf{M}_{\bar{g}\theta}\varphi(X). \tag{3}$$

Definición 2. El problema de decisión estadística, relacionado con el juego estadístico $(\mathscr{D}, \Theta, w)$, $(X, \mathbf{P}_\theta)$, se llama *problema invariante respecto al grupo G*, si la familia $\mathbf{P}_\theta$ es invariante respecto a G, y la función de pérdidas w es invariante respecto a G en el sentido siguiente: para cualesquiera $\delta \in D$, $g \in G$ existirá el único $\delta' \in D$ tal, que

$$w(\delta, \theta) = w(\delta', \bar{g}\theta) \quad \text{para todos} \quad \theta \in \Theta. \tag{4}$$

El valor δ', unívocamente definido respecto a g, lo designaremos por $g'\delta$.

Lema 1. *Las transformaciones g' del espacio D en sí, engendradas por el grupo G, forman el grupo G'.*

Demostración. Mostraremos que la población G' de todas las transformaciones g' está cerrada respecto a la composición y que además es válida la igualdad $g_2' g_1' = (g_2 g_1)'$.

En efecto,

$$w(\delta, \theta) = w(g_1'\delta, \bar{g}_1\theta) = w(g_2' g_1'\delta, \bar{g}_2\bar{g}_1\theta) = w((g_2 g_1)'\theta, \overline{(g_2 g_1)}\theta).$$

Como $\overline{(g_2 g_1)} = \bar{g}_2\bar{g}_1$, entonces, en virtud de la unicidad, $(g_2 g_1)' = g_2' g_1'$. El lema queda demostrado.

Así pues, con el principal grupo G de las transformaciones g del espacio $\mathscr{X}^n$ en sí, están relacionados otros dos grupos $\bar{G}$ y G' de transformaciones de los espacios Θ y D en sí. El empleo simultáneo de las tres transformaciones g, $\bar{g}$ y g' deja inalterable (invariante) el problema de decisión. Por eso es natural elegir tales reglas de decisión que no varíen al pasar de un problema de decisión equivalente a otro. En los §§ 2.18, 2.19 y 3.7 ya hemos analizado muy detalladamente la naturaleza de tal enfoque.

Definición 3. La función de decisión $\delta(X)$ del problema invariante de decisión se llama *invariante* si

$$\delta(gX) = g'\delta(X).$$

La *regla invariante randomizada* $\pi(X)$ se define como cualquier distribución concentrada en las reglas invariantes de decisión.

Ejemplos de utilización del principio de invariación se ofrecen en los §§ 2.18, 2.19 y 3.7 ya mencionados, donde hemos examinado las estimaciones equivariantes y los criterios invariantes. Es preciso señalar cierta peculiaridad de estos dos casos particulares desde el punto de vista del enfoque general.

En el *problema de estimación*, el grupo de transformación G' no se ha introducido en absoluto. En este caso, los conjuntos D y Θ coinciden, y desde el principio se suponía que $g'\delta = \bar{g}\delta$. Por eso hemos definido las estimaciones equivariantes con ayuda de la igualdad $\theta^*(gX) = \bar{g}\theta^*(X)$.

En la *teoría de verificación de hipótesis* se suponía que la transformación g' era igual a la transformación idéntica $g' = e$, por lo tanto, el criterio invariante π podía ser definido por la relación $\pi(gX) = \pi(X)$.

En este caso, para la invariación del problema de verificación de dos hipótesis $\{\theta \in \Theta_1\}$ y $\{\theta \in \Theta_2\}$ también es necesario suponer (véase (4)) que $g\Theta_i = \Theta_i$.

Precisamente debido a la existencia de cierta diferencia en estos dos enfoques se explica, en cierta medida, la utilización de dos términos diferentes: "equivariación" (para las estimaciones) e "invariación" (para la verificación de hipótesis) para designar las reglas de decisión invariantes. Adicionalmente a los ejemplos de problemas invariantes de decisión, examinados en los capítulos 2 y 3, citaremos uno más.

Ejemplo 1. Supongamos que $X \in \Phi_{\alpha,\sigma^2}$. Aquí Θ es el semiplano $\{\theta = (\alpha, \sigma): \sigma \geqslant 0\}$. Sea D la recta real R, y sea $w(\delta, \theta) = (\delta - \alpha)^2/\sigma^2$.

Examinemos el grupo G de transformaciones $g_{a,b}X = a + bX = (a + bx_1, \dots, a + bx_n)$, donde $b \neq 0$. La variable aleatoria $g_{a,b}X$ en $\mathscr{X}^n$ puede, evidentemente, considerarse como una muestra de $\Phi_{a+b\alpha,\, b^2\sigma^2}$. Por consiguiente, la familia Φ_{α,σ^2} es invariante respecto a G, si se pone $\bar{g}_{a,b}\theta = (a + b\alpha, |b|\sigma)$. La función de pérdidas será invariante si ponemos $g'_{a,b}\delta = a + b\delta$, puesto que

$$w(g'_{a,b}\delta, \bar{g}_{a,b}\theta) = \frac{(a + b\delta - a - b\alpha)^2}{b^2\sigma^2} = w(\delta, \theta).$$

Ahora bien, tenemos un problema invariante de decisión respecto a G. Las funciones invariantes de decisión $\delta(X)$: $\mathscr{X}^n \to R$ deben poseer la propiedad

$$\delta(a + bX) = \delta(g_{a,b}X) = g'_{a,b}\delta(X) = a + b\delta(X). \tag{5}$$

Seguidamente, no es difícil establecer que el problema de decisión sometido a examen también es invariante respecto al grupo F de todas las permutaciones f de las coordenadas del vector X; en este caso, $\bar{f}$ y f' serán dos transformaciones idénticas. Por eso, si exigimos, que la función $\delta(X)$ también sea una decisión invariante respecto a F, entonces también debe cumplirse

$$\delta(fX) = \delta(X). \tag{6}$$

Nótese que la clase de funciones que satisfacen (5) y (6) aún es bastante amplia: en ella entran, por ejemplo, todas las formas lineales

$$\delta(X) = \sum_{k=1}^{n} a_k x_{(k)}, \quad \sum_{k=1}^{n} a_k = 1,$$

donde $x_{(1)}, \dots, x_{(n)}$ es la serie variacional de la muestra X. Si utilizamos el principio de no desplazamiento, obtendremos una condición más para los coeficientes a_k:

$$\sum_{k=1}^{n} a_k \mathbf{M}_\theta(x_{(k)} - \alpha) = 0. \lhd$$

Al construir las decisiones invariantes óptimas en la teoría de estimación y en la teoría de verificación de las hipótesis estadísticas, desempeñan un papel muy importante los conceptos que, en cierto sentido, se asemejan uno a otro: el concepto de órbita en la teoría de estimaciones, y el concepto de invariante en la teoría de verificación de hipótesis. Recordemos que por órbita en el espacio Θ se entiende el conjunto $\{\bar{g}\theta_0, \bar{g} \in \bar{G}\}$, donde θ_0 es cierto punto de Θ. Con otras palabras, θ_1 y θ_2 pertenecen a una misma órbita, si existe $\bar{g} = \bar{C}$ tal, que $\theta_1 = \bar{g}\theta_2$.

Análogamente se pueden definir las órbitas en $\mathscr{X}^n$. Entonces son invariantes, por definición, las estadísticas constantes en las órbitas en $\mathscr{X}^n$.

El concepto de órbita también conserva su importancia en el caso general.

Lema 2. *La función de riesgo del problema invariante de decisión para una regla invariante de decisión, es constante en la órbita:*

$$W(\delta(\cdot), \theta) = W(\delta(\cdot), \bar{g}\theta)$$

para todos $\theta \in \Theta$, $\bar{g} \in \bar{G}$.

Demostración. En virtud de la invariación respectiva de la función de pérdidas, de la regla de decisión y de la familia $\mathbf{P}_\theta$ (véanse (3) y (4)), tenemos

$$W(\delta(\cdot), \theta) = \mathbf{M}_\theta w(\delta(X), \theta) = \mathbf{M}_\theta w(g'\delta(X), \bar{g}\theta) =$$
$$= \mathbf{M}_\theta w(\delta(gX), \bar{g}\theta) = \mathbf{M}_{\bar{g}\theta} w(\delta(X), \bar{g}\theta) = W(\delta(\cdot), \bar{g}\theta). \lhd$$

La constancia en la órbita de riesgo para las reglas de decisión invariantes randomizadas se deduce de su definición y del lema 2.

De este último resulta que en el caso de que todo el espacio Θ sea una órbita (es decir, $\Theta = \{\bar{g}\theta_0, \bar{g} \in \bar{G}\}$ para cualquier θ_0; esto tiene lugar, por ejemplo, para las transformaciones de desplazamiento), la regla invariante de decisión será una regla igualadora. Por eso, del lema 2 y de los teoremas 2.3, 2.5 obtenemos directamente la siguiente afirmación que establece una relación importante entre la invariación y el carácter minimax.

Teorema 2. *Supongamos que el espacio* Θ *es una órbita y que existe una distribución a priori* $\mathbf{Q}$ *para la cual la estrategia bayesiana* $\pi_Q(X)$ *es invariante. Entonces* $\pi_Q(X)$ *será una estrategia minimax.*

Del teorema 3.3 se desprende que tiene lugar la siguiente generalización del teorema 2.

Teorema 2A. *Supongamos que existe una distribución a priori* $\mathbf{Q}$ *concentrada en una de las órbitas, tal, que la estrategia* Θ_0 *bayesiana* $\pi_Q(X)$ *es invariante.*

Entonces, si para todos θ,

$$W(\pi_Q(\cdot), \theta) \leqslant W(\pi_Q(\cdot), \theta_0), \quad \theta_0 \in \Theta_0,$$

entonces $\pi_Q(X)$ *es minimax.*

Este criterio fue utilizado en el § 3.9.

§ 6. Estimaciones asintóticamente óptimas para una función de pérdidas arbitraria

Muchos de los resultados de las estimaciones asintóticamente óptimas (capítulo 2) y de los criterios asintóticamente óptimos (capítulo 3) admiten generalizaciones en la función de pérdidas, de forma muy general.

En este párrafo investigaremos los problemas de la teoría de estimación y supondremos que $w(\delta, \theta) = w(\delta - \theta)$.

Hagamos primeramente una observación general. En el capítulo 2 hemos visto que en el caso general ($X \in \mathbf{P}_\theta$, $\mathbf{P}_\theta$ satisface las condiciones (RR); véanse los §§ 2.24 y 2.28), todas las estimaciones racionales $\theta^* = \delta(X)$ del parámetro θ están "concentradas" en el entorno $1/\sqrt{n}$ del punto θ. Así, por ejemplo, para las estimaciones asintóticamente normales, $(\theta^* - \theta)\sqrt{n} \Subset \Phi_{0,\sigma^2(\theta)}$. De aquí se deduce que, para amplias suposiciones respecto a la función $w(t)$, el comportamiento asintótico del riesgo $\mathbf{M}_\theta w(\theta^* - \theta)$ será determinado por las propiedades de la función $w(t)$ en el entorno del punto $t = 0$. Si $w(t)$ es dos veces continuamente derivable en el cero, $w'' > 0$, entonces, para $t \to 0$,

$$w(t) = \frac{w''(0)}{2} t^2 + o(t^2). \tag{1}$$

Esto significa que en la región de valores de t (del orden de $1/\sqrt{n}$) que nos interesa, la función $w(t)$ se comportará igual que la función cuadrática de pérdidas $w_0(t) = ct^2$, cuando $c = w''(0)/2$, para la que han sido establecidos los resultados del capítulo 2. Si, además, $w(t) < e^{\alpha|t|^2}$, siendo bastante pequeño $\alpha > 0$ (véase el teorema 2.28.6), todos estos resultados mantendrán su validez, ya que su traslado al caso de la función $w(t)$ de forma (1), es cuestión de una técnica no complicada, completamente al alcance del lector.

En este párrafo examinaremos una generalización mucho más sustancial. Supondremos que la función de pérdidas $w(\delta, \theta)$ dependa de n y que la misma es representable en la forma

$$w(\delta, \theta) = w_n(\delta - \theta) = w(\sqrt{n}(\delta - \theta)), \tag{2}$$

donde la función $w(t) \geqslant 0$ está definida en todo el espacio R^k. Es evidente que en este caso serán esenciales los valores de $w(t)$ en toda la región de los valores de t.

Admitiremos que la función w en (2) satisface las condiciones siguientes:

1) *$w(t) \leqslant e^{c|t|}$ para cierto $c > 0$.*

Tal forma de condición 1) simplifica algo los cálculos. En efecto, todos los resultados conservarán su validez si exigimos que $w(t) \leqslant c_1 e^{\alpha|t|^2}$ cuando $\alpha > 0$ es bastante pequeño.

Posteriormente desempeñará un papel muy importante la función

$$V_{\sigma^2}(s) = \int w(s-u)e^{-\frac{1}{2}u\sigma^2u^T}du,$$

donde σ^2 es cierta matriz de segundos momentos, definida positivamente. La función $V_{\sigma^2}(s)$ puede interpretarse como

$$V_{\sigma^2}(s) = \frac{(2\pi)^{k/2}}{\sqrt{|\sigma^2|}}\mathbf{M}w(s-\xi), \quad \xi \in \Phi_{0,\sigma^{-2}}.$$

En vista de que

$$V_{\sigma^2}(s) = \int w(v)e^{-\frac{1}{2}(s-v)\sigma^2(s-v)^T}dv,$$

esta función será la función analítica de las variables s y σ^2.

También necesitaremos las condiciones:

2) *La función $V_{\sigma^2}(s)$ alcanza su valor mínimo respecto a s en un solo punto que designaremos por b_w.*

3) $b_w = 0$.

4) *La función $w(t)$ es continua.*

La condición 2) se cumplirá a ciencia cierta si $w(s) \neq$ const es una función convexa hacia abajo. En este caso $V_{\sigma^2}(s)$ será, evidentemente, también convexa y no contendrá partes "lineales" (o sea, la matriz de segundas derivadas será por doquier definida positivamente).

La condición 3) será cumplida si

$$V'_{\sigma^2}(0) = -\int uw(u)e^{-\frac{1}{2}u\sigma^2u^T}du = 0,$$

lo cual siempre tendrá lugar para las funciones simétricas $w(u) = w(-u)$.

El valor de b_w podría llamarse *desplazamiento* de la función de pérdida w. El mismo satisface la ecuación $V'_{\sigma^2}(b_w) = 0$. La condición 3) acerca de que $b_w = 0$ no es esencial y sólo simplifica la exposición, que el lector también puede extender fácilmente al caso de $b_w \neq 0$. Las modificaciones que en este caso tendrán lugar en los enunciados de los teoremas, serán ilustradas en la observación 2 correspondiente al teorema 1.

Recordemos ahora en qué se transformarán las definiciones de las estrategias óptimas expuestas en los §§ 2 y 3. La estimación θ^*_Q será *bayesiana* respecto a la distribución a priori $\mathbf{Q}$ con densidad q respecto a la medida de Lebesgue (y a la función de pérdidas w_n) si

$$\int W(\theta^*_Q, t)q(t)dt = \min_{\theta^*} \int W(\theta^*, t)q(t)dt, \qquad (3)$$

donde $W(\theta^*, t) = \mathbf{M}_t w_n(\theta^* - t)$. Aquí la integral del segundo miembro (3) puede escribirse en forma de la esperanza matemática incondicional $\mathbf{M}w_n(\theta^* - \theta)$, donde la promediación se toma respecto a la distribución con densidad $f_t(x)q(t)$.

La estimación $\bar{\theta}^*$ será *minimax* si para cualquier otra estimación θ^*,

$$\sup_t W(\bar{\theta}^*, t) \leqslant \sup_t W(\theta^*, t).$$

Lo dicho hace naturales las siguientes definiciones que son completamente análogas a las dadas en el § 2.11.

Definición 1. Llamaremos *asintóticamente bayesiana* la estimación θ^* si

$$\lim_{n\to\infty} \sup \, [\mathbf{M} w_n(\theta^* - \theta) - \mathbf{M} w_n(\theta_Q^* - \theta)] \leqslant 0, \tag{4}$$

donde θ_Q^* es la estimación bayesiana.

Definición 2. Llamaremos *asintóticamente minimax* la estimación θ_1^*, si para cualquier otra estimación θ^*,

$$\lim_{n\to\infty} \sup \left[\sup_{t\in\Theta_0} W(\theta_1^*, t) - \sup_{t\in\Theta_0} W(\theta^*, t)\right] \leqslant 0, \tag{5}$$

donde Θ_0 es cualquier subconjunto cerrado que se encuentra dentro de Θ.

Al estudiar las estimaciones asintóticamente óptimas en este párrafo, sólo utilizaremos los conceptos introducidos en las definiciones 1 y 2. Esto constituye cierta diferencia del capítulo 2, donde también estaban presentes las estimaciones asintóticamente eficientes. Aquí su ausencia se explica por el hecho de que para las funciones arbitrarias de pérdidas w no disponemos de desigualdades del tipo de Rao — Cramer para $\inf_{\theta^*\in K_0} W(\theta^*, \theta)$ (K_0 es la clase de estimaciones no desplazadas), con ayuda de la cual era posible, valiéndose del valor de $W(\theta^*, \theta)$, juzgar acerca de la calidad de θ^* y destacar, en particular, las estimaciones eficientes (y asintóticamente eficientes), o sea, las estimaciones uniformemente mejores en la clase K_0.

Las afirmaciones siguientes establecen que la estimación de verosimilitud máxima es, al igual que en las condiciones del capítulo 2, asintóticamente bayesiana y asintóticamente minimax. Además, obtendremos la frontera inferior *asintótica* para la función de riesgo al ser arbitraria la función de pérdidas w (la desigualdad de Rao — Cramer proporciona la frontera inferior *exacta*). En los tres teoremas ulteriores supondremos que se cumple la condición (RR).

Teorema 1. *Supongamos que* $X \in \mathbf{P}_\theta$, θ^* *es la e.v.m., y que* θ_Q^* *es una estimación bayesiana correspondiente a la función de pérdidas* w (véase (2)) *que satisface las condiciones* 1) — 3), *así como a la distribución a priori* $\mathbf{Q}$ *con una densidad* q *limitada respecto a la medida de Lebesgue. Entonces*

$$|\theta_Q^* - \hat{\theta}^*|\sqrt{n} \underset{P_\theta}{\to} 0, \tag{6}$$

$$(\theta_Q^* - \theta)\sqrt{n} \in \Phi_{0,I^{-1}(\theta)} \tag{7}$$

es uniforme respecto a $\theta \in \Theta_0$; Θ_0 *cualquier subconjunto cerrado, situado dentro de* Θ, *en el que* $q(\theta) > q_0 > 0$ *es continua.*

Si, además, la función w satisface la condición (4), *entonces*

$$\mathbf{M}w_n(\theta^*_Q - \theta) = \mathbf{M}w(\sqrt{n}(\theta^*_Q - \theta)) \to \mathbf{M}w(\eta_\theta) = \mathbf{M}\, \frac{\sqrt{I(0)}}{(2\pi)^{k/2}}\, V_{I(\theta)}(0), \tag{8}$$

donde $\eta_\theta \in \Phi_{0,I^{-1}(\theta)}$, $\theta \in \mathbf{Q}$; $\mathbf{M}$, *como antes, designa la esperanza matemática incondicional cuya densidad constituye* $f_t(x)\, q(t)$ $(X \in \mathbf{P}_\theta,\ \theta \in \mathbf{Q})$.

Observación 1. A la par con la convergencia (6) también se puede establecer una convergencia casi segura respecto a $\mathbf{P}_\theta$.

Observación 2. Si w es tal que el desplazamiento $b_w \neq 0$, la afirmación del teorema 1 quedará válida por completo, siempre que θ^*_Q en (6), (7) y (8) se sustituya por $\theta^*_Q - b_w/\sqrt{n}$. Ahora bien, b_w tiene sentido de *desplazamiento asintótico de la magnitud* $(\theta^*_Q - \theta)\sqrt{n}$.

Teorema 2. *Supongamos que la función w satisface las condiciones* 1) — 4). *Entonces, para cualquier estimación* θ^*,

$$\lim_{n\to\infty}\ \inf \sup_{t\in\Theta_\epsilon} \mathbf{M}_t w_n(\theta^* - t) \geqslant \sup_{t\in\Theta_\epsilon} \mathbf{M}w(\eta_t), \tag{9}$$

$$\eta_t \in \Phi_{0,I^{-1}(t)}.$$

Cualquier estimación θ^* *para la cual*

$$\mathbf{M}_t w_n(\theta^* - t) \to \mathbf{M}w(\eta_t) \tag{10}$$

uniformemente respecto a t, es asintóticamente minimax.

Teorema 3. *Supongamos que* $X \in \mathbf{P}_\Theta$ *y que la función w satisface las condiciones* 1) — 4). *Entonces, la estimación de verosimilitud máxima* $\hat{\theta}^*$ *es asintóticamente minimax y asintóticamente bayesiana para cualquier distribución a priori* $\mathbf{Q}$ *cuya densidad q es continuamente positiva en el punto* θ.

Todas estas afirmaciones son absolutamente análogas a las afirmaciones correspondientes del capítulo 2, ya que las mismas contribuyen a la verosimilitud de la suposición de que también para la función de pérdidas arbitraria w que satisface las condiciones 1) — 4), la e.v.m. es la mejor estimación asintóticamente uniforme en la clase de estimaciones asintóticamente no desplazadas (compárese con los §§ 2.25 y 2.28).

Demostración del teorema 1. En virtud del principio bayesiano, la estimación bayesiana se define como el valor θ^*_Q que posee la propiedad

$$\int w_n(\theta^*_Q - t)q(t/X)dt = \min_{u\in\Theta} \int w_n(u - t)q(t/X)dt =$$
$$= \min_{u\in\Theta} \int w(\sqrt{n}(u-\theta) - \sqrt{n}(t-\theta))\, \frac{q(t)f_t(X)}{\int q(v)f_v(X)dv}\, dt.$$

Esto significa que en calidad de $(\theta^*_Q - \theta)\sqrt{n} = u^*_Q$ se puede tomar cualquier

valor s con el cual se alcanza mín $U(s)$,

$$U(s) = \int w(s - v)q\left(\theta + \frac{v}{\sqrt{n}}\right)Z\left(\frac{v}{\sqrt{n}}\right)dv, \tag{11}$$

donde, como antes, $Z(t) = \dfrac{f_{\theta+t}(X)}{f_\theta(X)}$.

Necesitaremos las afirmaciones acerca del comportamiento asintótico de $U(s)$. En los §§ 2.28 y 2.29 hemos establecido (teorema 2.28.5) que, al cumplirse las condiciones (RR),

$$U(u^*) = e^{Y(u^*)}q(\hat{\theta}^*)(V_{I(\theta)}(0) + \varepsilon_n(X, \theta)), \tag{12}$$

donde $\varepsilon_n(X, \theta) \underset{P}{\to} 0$ uniformemente respecto a θ (aquí hemos sustituido $\dfrac{(2\pi)^{k/2}}{\sqrt{I(\theta)}}\mathbf{M}w(\xi)$ por $V_{I(\theta)}$, y $q(\theta)$ por $q(\hat{\theta}^*)$).

Nótese ahora que

$$\mathbf{P}(\sqrt{n}\,|\theta_Q^* - \hat{\theta}^*| \geqslant \varepsilon) = \mathbf{P}(|u_Q^* - u^*| \geqslant \varepsilon) \leqslant$$
$$\leqslant \mathbf{P}\left(\min_{|s-u^*|\geqslant\varepsilon} U(s) \leqslant U(u^*)\right). \tag{13}$$

En vista de que tenemos la representación asintótica para $U(u^*)$, aquí debemos estimar el valor de $U(s)$. De los teoremas 2.28.4 y 2.29.3 se deduce que para la sucesión arbitraria $\delta_n \to 0$, cuando $|v| < \delta_n\sqrt{n}$,

$$\ln Z\left(\frac{v}{\sqrt{n}}\right) = Y(u^*) - \frac{1}{2}(v - u^*)I(\theta)(v - u^*)^T(1 + \varepsilon_n(X, \theta, u)),$$

$|\varepsilon_n(X, \theta, u)| \leqslant \varepsilon_n^{(1)}(X, \theta) \underset{P_\theta}{\to} 0$ uniformemente respecto θ. Pero

$$U(s) \geqslant U_n(s) \equiv \int\limits_{|v-u^*|\leqslant\delta_n\sqrt{n}} w(s - v)q\left(\theta + \frac{v}{\sqrt{n}}\right)Z\left(\frac{v}{\sqrt{n}}\right)dv.$$

Examinemos el conjunto

$$A_n = \left\{\varepsilon_n^{(1)}(X, \theta) < \varrho, \inf_{\substack{|v-u^*|\leqslant\delta_n\sqrt{n}\\ \varrho>0,}} q\left(\theta + \frac{v}{\sqrt{n}}\right) > q(\hat{\theta}^*)(1 - \varrho)\right\},$$

que posee, evidentemente, la propiedad

$$\mathbf{P}_\theta(A_n) \to 1. \tag{14}$$

En este conjunto, uniformemente respecto a θ,

$$U_n(s) \geqslant (1 - \varrho)q(\hat{\theta}^*)e^{Y(u^*)} \times$$
$$\times \int\limits_{|v-u^*|\leqslant\delta_n\sqrt{n}} w(s - v)\exp\left\{-\frac{1}{2}(v - u^*)I(\theta)(v - u^*)^T(1 + \varrho)\right\}dv =$$
$$= (1 - \varrho)q(\hat{\theta}^*)e^{Y(u^*)}[V_{I(\theta)(1+\varrho)}(s - u^*) - r_n(s)], \tag{15}$$

donde, según la condición 1),

$$r_n(s) = \int\limits_{|v-u^*| \leqslant \delta_n\sqrt{n}} w(s-v) \exp\left\{-\frac{1}{2}(v-u^*)I(\theta)(v-u^*)^T \times \right.$$

$$\left. \times (1+\varrho)\right\} dv \leqslant e^{c\sqrt{n}d} \frac{(2\pi)^{k/2}}{\sqrt{|I(\theta)|c}} \mathbf{P}(|\eta| > \delta_n\sqrt{n}),$$

$$\eta \in \Phi_{0,I(\theta)(1+\varrho)},$$

donde d es el diámetro de la región Θ. Al igual que en el lema 2.23.1, es fácil convencerse de que

$$\mathbf{P}(|\eta| > \delta_n\sqrt{n}) \leqslant e^{-\alpha n\delta_n^2}, \quad \alpha > 0.$$

Eligiendo $\delta_n = n^{-1/9}$, obtenemos que, para todos los valores de s y con valores de n bastante grandes,

$$r_n(s) \leqslant e^{-n^{3/4}}. \tag{16}$$

Ahora utilicemos las condiciones 2) y 3) en virtud de las cuales

$$\min_{|s-u^*|>\varepsilon} V_{I(\theta)}(s-u^*) \geqslant V_{I(\theta)}(0) + 4\tau, \quad \tau = \tau(\varepsilon) > 0.$$

En virtud de las propiedades analíticas de $V_{\sigma^2}(s)$ obtendremos que, para valores de ϱ bastante pequeños,

$$\min_{|s-u^*|>\varepsilon} V_{I(\theta)(1+\varrho)}(s-u^*) \geqslant V_{I(\theta)}(0) + 3\tau,$$

y en virtud de (15) y (16), para $X \in A_n$ y para valores de n bastante grandes,

$$\min_{|s-u^*|>\varepsilon} U_n(s) \geqslant (1-\varrho)q(\hat{\theta}^*)e^{Y(u^*)}[V_{I(\theta)}(0) + 2\tau].$$

Utilizando (12) y (13), definitivamente obtenemos

$$\mathbf{P}_\theta(\sqrt{n}|\theta_Q^* - \hat{\theta}^*| \geqslant \varepsilon) \leqslant \mathbf{P}_\theta(\min_{|s-u^*|>\varepsilon} U_n(s) \leqslant U(u^*)) \leqslant$$

$$\leqslant \mathbf{P}_\theta(X \notin A_n) + \mathbf{P}_\theta((1-\varrho)[V_{I(\theta)}(0) + 2\tau] \leqslant V_{I(\theta)}(0) + \varepsilon_n(X, \Theta)).$$

Eligiendo adicionalmente ϱ, de tal modo que su valor sea tan pequeño que contribuya al cumplimiento de $(1-\varrho)2\tau - V_{I(\theta)}(0) \geqslant \tau$, obtendremos

$$\mathbf{P}(\sqrt{n}|\theta_Q^* - \hat{\theta}^*| \geqslant \varepsilon) \leqslant \mathbf{P}_\theta(X \notin A_n) + \mathbf{P}_\theta(\varepsilon_n(X, \theta) > \tau) \to 0$$

cuando $n \to \infty$. En virtud de (12) y (14), la afirmación (6) queda demostrada.

De (6) y de los teoremas del § 2.29 se desprende (7). Demostremos ahora la relación (8). En virtud de (7) y de la propiedad (4),

$$w(\sqrt{n}(\theta_Q^* - \theta)) \Rightarrow w(\eta_\theta), \quad \eta_\theta \in \Phi_{0,I^{-1}(\theta)}.$$

Según el lema de Fatou,

$$\lim_{n\to\infty} \inf \mathbf{M}_t w(\sqrt{n}(\theta_Q^* - t)) \geqslant \mathbf{M}w(\eta_t),$$

$$\lim_{n\to\infty} \inf \mathbf{M}w(\sqrt{n}(\theta_Q^* - \theta)) \geqslant \int q(t)\mathbf{M}w(\eta_t)dt = \mathbf{M}w(\eta_t) = \mathbf{M}w(\eta_\theta).$$

Por otro lado, según la definición de θ_Q^*,

$$\mathbf{M}w(\sqrt{n}(\theta_Q^* - (\theta)) \leqslant \mathbf{M}w(\sqrt{n}(\hat{\theta}^* - \theta)) \to \mathbf{M}w(\eta_\theta).$$

La última relación se deduce de la convergencia uniforme $\mathbf{M}_t w(\sqrt{n}(\theta^* - t)) \to \mathbf{M}w(\eta_t)$ demostrada en el § 2.29. El teorema queda demostrado.

Demostración del teorema 2. Tomemos la distribución $\mathbf{Q}$ concentrada en Θ_0, con una densidad limitada $q(t) > 0$ para $t \in \Theta_0$, y sea θ_Q^* la estimación bayesiana correspondiente a $\mathbf{Q}$. Entonces, para cualquier estimación θ^*,

$$\sup_{t\in\Theta_0} \mathbf{M}_\theta w_n(\theta^* - t) \geqslant \int_{\Theta_0} \mathbf{M}_t w_n(\theta^* - t)q(t)dt \geqslant$$

$$\geqslant \int_{\Theta_0} \mathbf{M}_t w_n(\theta_Q^* - t)q(t)dt = \mathbf{M}w_n(\theta_Q^* - \theta).$$

Según el lema de Fatou, en virtud de (8),

$$\lim_{n\to\infty} \inf \sup_{t\in\Theta_0} \mathbf{M}_t w_n(\theta^* - t) \geqslant \lim_{n\to\infty} \inf \mathbf{M}w_n(\theta_Q^* - \theta) \geqslant \mathbf{M}w(\eta_\theta) =$$

$$= \int_{\Theta_0} \mathbf{M}w(\eta_t)q(t)dt.$$

Como la función $\mathbf{M}w(\eta_t) = \dfrac{\sqrt{I(t)}}{(2\pi)^{k/2}} V_{I(t)}(0)$ es continua respecto a t, entonces, eligiendo $q(t)$, podemos conseguir que la integral

$$\int_{\Theta_0} \sqrt{I(t)}\, V_{I(t)}(0)q(t)dt$$

se asemeje tanto como se quiera a $\sup_{t\in\Theta_0} \sqrt{I(t)}\, V_{I(t)}(0) = \sup_{t\in\Theta_0} \mathbf{M}w(\eta_t)$. Esto demuestra (9).

Ahora supongamos que la estimación θ_1^* posee la propiedad (10), y que θ^* es cualquier otra estimación. Entonces, en virtud de (9) y de la convergencia uniforme (10),

$$\lim_{n\to\infty} \sup \left[\sup_{t\in\Theta_0} \mathbf{M}_t w_n(\theta_1^* - t) - \sup_{t\in\Theta_0} \mathbf{M}_t w_n(\theta^* - t)\right] \leqslant$$

$$\leqslant \sup_{t\in\Theta_0} \lim_{n\to\infty} \mathbf{M}_t w_n(\theta_1^* - t) - \sup_{t\in\Theta_0} \mathbf{M}_w(\eta_t) = 0.$$

La desigualdad (5) de definición del carácter asintóticamente minimax, y junto con ella el teorema 2, quedan demostrados.

Demostración del teorema 3. El carácter asintóticamente minimax de $\hat{\theta}^*$ se desprende del hecho de que para la e.v.m. $\hat{\theta}^*$, según el teorema 2.29.4, es válida (10).

El carácter asintóticamente bayesiano de $\hat{\theta}^*$ se deduce del hecho de que para $\theta^* = \hat{\theta}^*$ se cumple (4), ya que para $\hat{\theta}^*$ tiene lugar la convergencia uniforme (10) y, por lo tanto,

$$\lim_{n\to\infty} \mathbf{M}w_n(\hat{\theta}^* - \theta) = \lim_{n\to\infty} \int \mathbf{M}_t w_n(\hat{\theta}^* - \theta) q(t)dt =$$
$$= \mathbf{M}w(\eta_\theta) = \lim_{n\to\infty} \mathbf{M}w_n(\theta_Q^* - \theta).$$

La última igualdad resulta de (8). El teorema queda demostrado.

La afirmación del teorema 1 puede ser reforzada si se exige adicionalmente que la función $w(t)$ aumente con bastante rapidez. Para esto, designemos $w_N = \min_{|t|>N} w(t)$ y $W_M = \max_{|t|\leqslant M} w(t)$ y examinemos la condición

5) *Existe* $\gamma < 1$ *tal, que* $w_N > 2W_{\gamma N}$ *para todos los valores de* N *bastante grandes.*

Si cuando $|t| \to \infty$, $w(t)$ crece como función potencial o exponencial, entonces se cumple la condición 5).

Teorema 4. *Si se cumplen las condiciones* 1) *y* 5) *cuando* $q(t) > q_0 > 0$ *en el conjunto cerrado* Θ_0, *y cuando* $q(t) \leqslant q_m < \infty$, *entonces, para ciertos valores de* $c < \infty$ *y de* $\alpha > 0$ *que no dependen de* t,

$$\mathbf{P}t(\sqrt{n}(\theta_Q^* - t) > N) \leqslant ce^{-\alpha N^2}, \quad t \in \Theta_0.$$

De aquí y del teorema 1 se deduce que para cualquier función continua $v(t)$ tal, que $|v(t)| \leqslant e^{-\alpha N^2/2}$, es válida

$$\mathbf{M}_t v(\sqrt{n}(\theta_Q^* - t)) \to \mathbf{M}v(\eta_t), \quad t \in \Theta_0.$$

Designemos

$$u(r) = \int_{|v|\geqslant r} w(-v)q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv$$

(ésta es la parte de la integral $U(0)$ que se encuentra en la región $|v| \geqslant r$. Para demostrar el teorema 4 necesitaremos el

Lema 1. *Si* $w(t)$ *satisface la condición* 1), *y* $q_m = \max_u q(u) < \infty$, *entonces, para ciertos* $\beta > 0$ *y* $a < \infty$ *que no dependen de* θ, *así como para todos* $0 < \delta < 1$,

$$\mathbf{P}_\theta(u(r) > \delta) \leqslant \frac{a}{\delta} e^{-\beta r^2}.$$

Esta desigualdad quedará válida para $w(t) \equiv 1$.

Demostración. Tenemos

$$\mathbf{P}_\theta(u(r) > \delta) \leqslant \mathbf{P}_\theta\left(\sup_{|v|\geqslant r} Z\left(\frac{v}{\sqrt{n}}\right) > 1 + \mathbf{P}_\theta(u(r) > \delta, \sup_{|v|\geqslant r} Z\left(\frac{v}{\sqrt{n}}\right) \leqslant 1\right).$$

La estimación del primer sumando se da en el teorema 2.23.2, en virtud de la cual este sumando no pasa de $c_1 e^{-r^2\beta}$, $\beta > 0$. El segundo sumando no supera

$$\mathbf{P}_\theta \left(\int_{|v| \geq r} w(-v) q\left(\theta + \frac{v}{\sqrt{n}}\right) Z^{1/2}\left(\frac{v}{\sqrt{n}}\right) dv > \delta \right). \tag{17}$$

Como, en virtud del teorema 2.23.1,

$$\mathbf{M}_\theta Z^{1/2}\left(\frac{v}{\sqrt{n}}\right) \leqslant e^{-2|v|^2\beta}, \ \beta > 0,$$

la esperanza matemática de la integral en (17) no superará (véase el lema 2.23.1.)

$$q_m \int_{|v| \geq r} e^{c|v|} e^{-2|v|^2\beta} dv \leqslant c_2 e^{-r^2\beta}.$$

Por eso, en virtud de la desigualdad de Chébishev, la probabilidad (17) no supera $c_2 e^{-r^2\beta}/\delta$. El lema queda demostrado.

Designemos por $u_1 r$ el valor de la integral $u(r)$ cuando $w(t) \equiv 1$:

$$u_1(r) = \int_{|v| \geq r} q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv.$$

Lema 2. *Si $q(\theta) > 0$ en el conjunto cerrado Θ_0, entonces, con cierto $b < \infty$ que no depende de θ, para cualquer $\varepsilon > 0$ y para todos los valores de n bastante grandes,*

$$\mathbf{P}_\theta(u_1(0) < \varepsilon) \leqslant b\varepsilon^2, \quad \theta \in \Theta_0.$$

Demostración. Para todos los valores de n bastante grandes tenemos

$$u_1(0) \geqslant \int_{|v| \leqslant 1} q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv \geqslant$$
$$\geqslant q_0 \int_{|v| \leqslant 1} \exp\left\{L\left(X, \theta + \frac{v}{\sqrt{n}}\right) - L(X, \theta)\right\} dv =$$
$$= q_0 \int_{|v| \leqslant 1} \exp\left\{(v, \zeta_n) + \frac{1}{2} v\gamma_n v^T\right\} dv,$$

donde

$$q_0 = \min_{t \in \Theta_0} q(t) > 0, \ \zeta_n = \frac{1}{\sqrt{n}} L'(X, t), \ \gamma_n = \frac{1}{n} \| L_{ij}''(X, \tilde{\theta}) \|,$$

$\tilde{\theta} = \theta + \varrho v n^{-1/2}$, $|\varrho| \leqslant 1$. (Aquí L' es el vector de las derivadas de la función logarítmica de verosimilitud; L_{ij}'', las derivadas parciales de segundo orden.) En vista de que $|v, \zeta_n| \leqslant |v||\zeta_n|$ y como, en virtud de las condiciones (RR),

$$|v\gamma_n v^T| \leqslant \frac{1}{n} \sum_{i=1}^{n} l(x_i) \sum_{i,j=1}^{k} |v_i v_j| \leqslant \frac{k|v|^2}{n} L_n,$$

donde $L_n = \sum_{i=1}^{n} l(x_i)$, entonces, en el conjunto $A = \{|\zeta_n| \leqslant 1/\varepsilon, L_n \leqslant n/\varepsilon^2 k\}$ es válida

$$u_1(0) \geqslant q_0 \int_{|v| \leqslant 1} \exp\left\{-\frac{|v|}{\varepsilon} - \frac{|v|^2}{2\varepsilon^2}\right\} dv \geqslant q_0 \varepsilon \int_{|s| \leqslant \varepsilon^{-1}} \exp\left\{-|s| - \frac{|s|^2}{2}\right\} ds \geqslant c_1 \varepsilon.$$

Esto quiere decir que tiene lugar el encaje $\{u_1(0) < c_1\varepsilon\} \subset \bar{A}$. Como

$$\mathbf{P}_\theta(\bar{A}) \leqslant \mathbf{P}_\theta(|\zeta_n| \geqslant \varepsilon^{-1}) + \mathbf{P}_\theta\left(L_n > \frac{n}{\varepsilon^2 K}\right) \leqslant \varepsilon^2 \mathbf{M}_\theta |\zeta_n|^2 + \frac{\varepsilon^2 k}{n} \mathbf{M}_\theta L_n,$$

$$\mathbf{M}_\theta |\zeta_n|^2 = \sum_{i=1}^{k} I_{ii}(\theta), \quad \mathbf{M}_\theta L_n = n\mathbf{M}_\theta l(x_1),$$

entonces

$$\mathbf{P}_\theta(\bar{A}) \leqslant c_2\varepsilon^2.$$

El lema queda demostrado.

Demostración del teorema 4. Designemos por M_n el conjunto de puntos s en los cuales se alcanza $\min U(s)$ (o sea, el conjunto de puntos $(\theta_Q^* - \theta)\sqrt{n}$; véase (11)) *). Entonces,

$$\{M_n \subset D\} = \left\{\min_{s\in D} U(s) < \min_{s\notin D} U(s)\right\}. \tag{18}$$

Por consiguiente,

$$\{\sqrt{n}|\theta_Q^* - \theta| > 2N\} = \left\{\min_{|s|>2N} U(s) < \min_{|s|\leqslant 2N} U(s)\right\} \subset \left\{\min_{|s|\geqslant 2N} U(s) < U(0)\right\}.$$

Aquí

$$\min_{|s|\geqslant 2N} U(s) \geqslant w_n \int_{|u|<N} q\left(\theta + \frac{u}{\sqrt{n}}\right) z\left(\frac{u}{\sqrt{n}}\right) du = w_N(u_1(0) - u_1(N)),$$

$$w_N = \min_{\substack{|s|>2N \\ |u|<N}} w(s-u) = \min_{|t|>N} w(t).$$

Seguidamente,

$$U(0) = \int w(-u) q\left(\theta + \frac{u}{\sqrt{n}}\right) z\left(\frac{n}{\sqrt{n}}\right) du \leqslant (u_1(0) - u_1(M) W_M + u(M),$$

donde $W_M = \max_{|t|\leqslant M} w(t)$,

De aquí obtenemos

$$\{\sqrt{n}|\theta_Q^* - \theta| > 2N\} \subset \{w_N(u_1(0) - u_1(N)) < w_M(u_1(0) - u_1(M)) + u(M)\} \subset$$
$$\subset \left\{\left(\frac{w_N}{W_M} - 1\right) u_1(0) < \frac{u(M)}{W_M} + \frac{u_1(N) w_N}{W_M} + u_1(M)\right\}.$$

En virtud de la condición 5) escojamos $M = \gamma N$, $\gamma < 1$ de modo que $w_n > 2W_M$ para todos los valores de N bastante grandes. Además, hagamos uso de las desigualdades $W_M > 2$ (para valores de M bastante grandes) $w_n < w(N) < e^{cN}$. Entonces es evidente que

$$\{\sqrt{n}|\theta_Q^* - \theta| > 2N\} \subset \{u_1(0) < u(\gamma N) + u_1(N)e^{cN}\}. \tag{19}$$

En virtud del lema 1 hallamos

$$\mathbf{P}_\theta\left(u(\gamma N) > \frac{1}{2} e^{-\alpha N^2}\right) \leqslant 2ae^{-\beta N^2\gamma^2 + \alpha N^2},$$

$$\mathbf{P}_\theta\left(u_1(N) > \frac{1}{2} e^{-cN - \alpha N^2}\right) \leqslant 2ae^{-\beta N^2 + \alpha N^2 + cN}.$$

---

*) En vez de M_n se podría examinar, por ejemplo, el menor punto (según la norma) en el que se alcanza $\min U(s)$.

Escogiendo $\alpha < \frac{1}{2}\beta\gamma^2$ obtenemos que, para valores de N bastante grandes, de (19) resulta

$$\mathbf{P}_\theta(\sqrt{n}\,|\theta_Q^* - \theta| > 2N) \leqslant 4ae^{-\alpha N^2} + \mathbf{P}_\theta(u_1(0) < e^{-\alpha N^2}).$$

Sólo nos queda hacer uso del lema 2, en virtud del cual

$$\mathbf{P}_\theta(u_1(0) < e^{-\alpha N^2}) \leqslant be^{-2\alpha N^2}.$$

El teorema queda demostrado.

§ 7. Criterios estadísticos óptimos para una función de pérdidas arbitraria. Criterio de la relación de verosimilitud como decisión asintóticamente bayesiana

1. Propiedades de optimización de los criterios estadísticos para una función de pérdidas arbitraria. En los párrafos precedentes hemos visto que muchos resultados principales de la teoría de estimación conservan su validez cualitativa al pasar a problemas más generales de la decisión estadística con pérdidas $w(\delta, \theta)$, $\delta \in D \subset R^k$, $\theta \in \Theta \subset R^k$, distintas de las cuadráticas.

El mismo cuadro se observa también en la teoría de verificación de las hipótesis. En el § 4 hemos visto que las reglas de decisión óptimas para los juegos con conjuntos finitos D y Θ y con función de pérdidas arbitraria, tienen la misma forma que los criterios óptimos para verificar un número finito de hipótesis simples, examinados en el § 3.1. Los resultados de los §§ 3.5—3.7, 3.9, 3.11, 3.13—3.15 también conservarán, en lo fundamental, su validez. En particular, los teoremas de los c.u.m.p., enunciados en los §§ 3.5—3.7, se transformarán en afirmaciones de las estrategias uniformemente mejores en los juegos estadísticos correspondientes ($\Theta \subset R^k$, $D = \{\delta_1, \delta_2\}$ es bipuntual), en los cuales, sin embargo, la función de pérdidas $w(\delta_i, \theta) = w_i(\theta)$, $w_i(\theta) = 0$ para $\theta \in \Theta_i$, $i = 1, 2$ ya no será obligatoriamente estadística ($w_i(\theta) = 1$ para $\theta \notin \Theta_i$), sino que tan sólo satisfará ciertas condiciones muy generales (por ejemplo, las propiedades de crecimiento monótono de $w_i(\theta)$ al alejarse θ de Θ_i). El papel de las clases K_ε, en las que hemos buscado los c.u.m.p., lo desempeñarán las clases de funciones de decisión $\pi(X)$, con valor máximo fijo ε de las "pérdidas de primer género":

$$\varepsilon = \sup_{\theta \in \Theta_1} W(\pi(\cdot), \theta) = \sup_{\theta \in \Theta_1} w_2(\theta)\mathbf{M}_\theta\pi(X, \delta_2). \tag{1}$$

Se minimizará el valor de las "pérdidas de segundo género":

$$W(\pi(\cdot), \theta) = w_1(\theta)\mathbf{M}_\theta\pi(X, \delta_1) \text{ para } \theta \in \Theta_2. \tag{2}$$

Aquí $\pi(X, \delta_i)$ significa la probabilidad de tomar la decisión δ_i a base del criterio π. Para abreviar la notación, pongamos, siguiendo el capítulo 3, $\pi(X, \delta_2) = \pi(X)$, así que $\pi(X, \delta_1) = 1 - \pi(X)$. La designación del criterio y del número $\pi(X, \delta_2)$ con ayuda de un solo símbolo $\pi(X)$ es cómoda y, como hemos visto antes, no produce equivocaciones.

En (1) y (2) se buscan los extremos de las expresiones que se distinguen de las expresiones correspondientes para las funciones estadísticas de pérdidas, tan sólo por los factores que no dependen de $\pi(X)$. Si estos factores poseen la propiedad natural de monotonía, entonces, al pasar al problema definido por (1) y (2), la exposición de los §§ 3.5—3.7, 3.9, 3.11 no variará considerablemente.

De hecho, también variarán poco los resultados de carácter asintótico en los §§ 3.13—3.15. En este párrafo examinaremos más detalladamente la generalización para el caso de una función de pérdidas arbitraria de los resultados del § 3.13 y nos convenceremos de que esta generalización realmente no exige ningunos esfuerzos adicionales.

2. C.r.v. como criterio asintóticamente bayesiano. Examinemos el juego estadístico $(\mathscr{D}, \Theta, W)$ en el que Θ es continual y constituye un conjunto compacto convexo en R^k, mientras que el conjunto D de estrategias del estadista es bipuntual: $D = \{\delta_1, \delta_2\}$. La función de pérdidas $w(\delta, \theta)$ tiene la forma siguiente:

$$w(\delta_1, \theta) = \begin{cases} 0, & \theta = \theta_1, \\ w_1(\theta), & \theta \neq \theta_1, \end{cases}$$
$$w(\delta_2, \theta) = \begin{cases} w_2, & \theta = \theta_1, \\ 0, & \theta \neq \theta_1, \end{cases}$$

donde θ_1 es un punto interior fijo de Θ. Cuando $w_2 = w_1(\theta) \equiv 1$ esto corresponde al problema de verificación de la hipótesis simple $H_1 = \{\theta = \theta_1\}$ frente a la alternativa adicional $H_2 = \{\theta \neq \theta_1\}$.

Para hallar, utilizando el principio bayesiano, la forma de decisión bayesiana, examinemos el juego corriente (D, Θ, w) y supongamos que en Θ se da una distribución $\mathbf{Q}$ tal, que $q = \mathbf{Q}(\{\theta_1\}) > 0$ (planteamiento bayesiano completo del problema). Designemos $\mathbf{Q}_2 = \dfrac{\mathbf{Q} - q\mathbf{I}_{\theta_1}}{1 - q}$, donde $\mathbf{I}_\theta$ es una distribución degenerada concentrada en el punto θ. Entonces

$$\tilde{w}(\delta_1, \mathbf{Q}) = (1 - q) \int w_1(t)\mathbf{Q}_2(dt), \quad \tilde{w}(\delta_2, \mathbf{Q}) = qw_2.$$

Esto quiere decir que la estrategia bayesiana $\pi_Q(\delta_2) = 1$ si

$$(1 - q) \int w_1(t)\mathbf{Q}_2(dt) > qw_2, \tag{3}$$

y $\pi_Q(\delta_1) = 1$ si tiene lugar la desigualdad inversa. La relación (3) puede escribirse en la forma

$$\int w(t)\mathbf{Q}(dt) > 0,$$

donde

$$w(t) = \begin{cases} w_1(t) & \text{para} \quad t \neq \theta_1, \\ -w_2 & \text{para} \quad t = \theta_1. \end{cases}$$

En virtud del principio bayesiano, la regla bayesiana de decisión $\pi_Q(X)$ tiene la forma $\pi_Q(X) = 1$ si

$$\int w(t)\mathbf{Q}_X(dt) > 0,$$

donde $\mathbf{Q}_X$ es la distribución a posteriori. Supongamos que $\lambda(dt) = dt$ para $t \neq \theta_1$, $\lambda(\{\theta_1\}) = 1$, y que la distribución $\mathbf{Q}_2$ tiene una densidad $q_2(t)$ respecto a la medida de Lebesgue. Entonces, la distribución $\mathbf{Q}$ tendrá una densidad $q(t)$ respecto a λ, igual a $(1 - q)q_2(t)$ para $t \neq \theta_1$, e igual a $q(t) = q$ para $t = \theta_1$. Esto significa que la densidad a posteriori respecto a la medida λ será igual a

$$q(t/X) = \frac{f_t(X)q(t)}{f(X)},$$

$$f(X) = \int f_u(X)q(u)\lambda(du).$$

Por consiguiente, la regla bayesiana de decisión $\pi_Q(X)$ tiene la forma $\pi_Q(X) = 1$ si

$$(1 - q)\int w_1(t)f_t(X)q_2(t)dt > w_2 q f_{\theta_1}(X). \tag{4}$$

El riesgo de esta regla es igual a

$$\bar{W}(\pi_Q(\cdot), \mathbf{Q}) = q w_2 \mathbf{P}_{\theta_1}(\pi_Q(X) = 1) + \\ + (1 - q)\int w_1(u)q_2(u)\mathbf{P}_u(\pi_Q(X) = 0)du.$$

Comparando estas relaciones con el contenido del § 3.13, vemos que la región (4) de toma de decisión δ_2 tiene aquí la misma forma que la región $\Omega(c)$ en (3.13.3) cuando $c = w_2 q/(1 - q)$ y cuando la función $q(t)$ en (3.13.3) se sustituye por $w_1(t)q_2(t)$. En otros términos,

$$\pi_Q(X) = \begin{cases} 1, & \text{si} \quad r_{Q_2}(X) > c, \\ \gamma, & \text{si} \quad r_{Q_2}(X) = c, \\ 0, & \text{si} \quad r_{Q_2}(X) < c, \end{cases} \tag{5}$$

donde

$$r_{Q_2}(X) = \frac{\int w_1(t)q_2(t)f_t(X)dt}{f_{\theta_1}(X)}, \quad c = \frac{w_2 q}{1 - q}.$$

Luego, siguiendo los razonamientos del § 3.13, podemos proceder del modo siguiente. De la población de reglas bayesianas (5) es necesario, modificando el número q, elegir tal decisión $\pi_Q(X)$, que tenga un valor fijo de "pérdidas de primer género":

$$w_2[\mathbf{P}_{\theta_1}(\pi_Q(X) = 1) + \gamma\mathbf{P}_{\theta_1}(\pi_Q(X) = \gamma)] = \alpha.$$

Entonces, entre todas las reglas $\pi(X)$, para las cuales

$$\alpha_1(\pi) = w_2 \mathbf{M}_{\theta_1}\pi(X) \leqslant \alpha, \tag{6}$$

la decisión $\pi_Q(X)$ minimizará las "pérdidas de segundo género" iguales a

$$\alpha_2(\pi) = \int w_1(u)q_2(u)\mathbf{M}_u(1 - \pi(X))du. \qquad (7)$$

Esto es la consecuencia directa del carácter bayesiano de la decisión π_Q. La comparación de los valores (6) y (7) con las magnitudes de las probabilidades de los errores de primero y segundo géneros (3.13.4.) muestra que otra vez se trata de distinciones no esenciales, la principal de las cuales consiste en que la función $q(u)$ en (3.13.4.) se sustituye por la función $w_1(u)q_2(u)$. Los números c y γ en (5) se determinan por α.

Lo dicho nos permite, siguiendo exactamente los razonamientos del § 3.13, enunciar las siguientes definiciones y afirmaciones.

Definición 1. La regla de decisión $\pi(X)$ pertenece a la clase $\tilde{\mathbf{K}}_\varepsilon$ (su nivel asintótico es $1 - \varepsilon$) si

$$\lim_{n\to\infty} \sup \mathbf{M}_{\theta_1}\pi(X) \leqslant \varepsilon.$$

Esta definición, de hecho, no se diferencia en nada de la definición 3.13.1.

Mostremos ahora que, eligiendo q, podemos tratar de que $\pi_Q \in \tilde{\mathbf{K}}_\varepsilon$. Pongamos

$$r_{Q_1}(X) = \frac{\int w_1(t)q_2(t)f_t(X)dt}{f_{\theta_1}(X)} = \left(\frac{2\pi}{n}\right)^{k/2} \frac{w_1(\theta_1)q_2(\theta_1)}{\sqrt{|I|}} e^{T(X)},$$

donde $I = I(\theta_1)$ es la matriz de información de Fisher en el punto θ_1. Supongamos, seguidamente, que se cumplen las condiciones (RR), que θ_1 es un punto interior en Θ, y que la función $w_1(t)\cdot q_2(t)$ es continua y positiva en el punto θ_1,

$$c = \left(\frac{2\pi}{n}\right)^{k/2} \frac{w_1(\theta_1)q_2(\theta_1)}{\sqrt{|I|}} e^z. \qquad (8)$$

Entonces, en viritud del lema 3.13.1., para la función $p_t(c) = \mathbf{P}_t(r_{Q_1}(X) > c)$ obtenemos

$$p_{\theta_1}(c) = \mathbf{P}_{\theta_1}(T(X) > z) \to \mathbf{H}_k((2z, \infty)).$$

Por consiguiente, poniendo $q = c/(c + w_2)$, donde c está definida en (8), $z = h_\varepsilon/2$, h_ε es una cuantila de orden $1 - \varepsilon$ de la distribución χ^2 con k grados de libertad, obtenemos

$$\lim_{n\to\infty} p_{\theta_1}\left(\frac{w_2 q}{1 - q}\right) = \varepsilon$$

y, por lo tanto, $\pi_Q(X) \in \tilde{\mathbf{K}}$.

Definición 2. Para una distribución a priori dada $\mathbf{Q}$, la regla de decisión $\pi(X)$ se llama *asintóticamente bayesiana* en $\bar{\mathbf{K}}_\varepsilon$ si $\pi_Q \in \bar{\mathbf{K}}_\varepsilon$,

$$\limsup_{n\to\infty} \frac{\alpha_2(\pi)}{\alpha_2(\pi_Q)} = 1.$$

Teorema 1. *Supongamos que se cumplen las condiciones (RR) y que* θ_1 *es un punto interior en* θ. *Entonces, en* $\bar{\mathbf{K}}_\varepsilon$ *existe una regla de decisión asintóticamente bayesiana* $\hat{\pi}(X)$ *que es la misma para cualesquiera distribuciones* $\mathbf{Q}_2$ *y para cualesquiera funciones* $w_1(t)$ *tales, que la función* $w_1(t)q_2(t)$ *es continua y positiva en el punto* θ_1 *y está limitada en* Θ. *El criterio* $\hat{\pi}$ *es definido por la relación*

$$\hat{\pi}(X) = 1 \quad \text{si} \quad \frac{f_{\theta^*}(X)}{f_{\theta_1}(X)} > e^{h_\varepsilon/2}. \tag{9}$$

El teorema se demuestra exactamente igual que el teorema 3.13.1, con una precisión de hasta la sustitución de la función $q(t)$ por $w_1(t)q_2(t)$. El teorema 3.13.1 también permite hallar el valor de las "pérdidas de segundo género" (véase (7)) del criterio $\hat{\pi}$.

El criterio (9) no es otra cosa sino el criterio de relación de verosimilitud.

§ 8. Soluciones asintóticamente óptimas para una función de pérdidas arbitrarias en el caso de hipótesis semejantes

En este párrafo examinaremos la generalización de los resultados del § 3.14 para el caso de una función de pérdidas arbitrarias. Esta generalización será más sustancial que en el párrafo anterior, ya que las funciones de pérdidas dependerán de n (compárese con el § 6).

Supongamos que $(\mathscr{D}, \Theta, W)$ es un juego estadístico en el que $\Theta \subset R^k$, el conjunto $D = \{\delta_1, \delta_2\}$ es bipuntual y $w(\delta_i, \theta) = w_i(\theta)$, donde $w_i(\theta) = 0$ cuando $\theta \in \Theta_i$, $i = 1, 2$, y la intersección $\Theta_1 \cap \Theta_2$ está vacía.

Si $w_i(\theta) = 1$ cuando $\theta \notin \Theta_i$, obtendremos el problema de verificación de las hipótesis $H_i = \{\theta \in \Theta_i\}$, $i = 1, 2$.

Determinemos la estrategia bayesiana para el juego (D, Θ, w). Sean $\mathbf{Q}_i$ las distribuciones en Θ_i,

$$\mathbf{Q} = q_1\mathbf{Q}_1 + q_2\mathbf{Q}_2, \quad q_1 + q_2 = 1.$$

Entonces es evidente que $\tilde{w}(\delta_i, \mathbf{Q}) = \int w_i(t)\mathbf{Q}(dt)$ y $\pi_Q(\delta_2) = 1$ si

$$\int w_2(t)\mathbf{Q}(dt) < \int w_1(t)\mathbf{Q}(dt),$$

o bien

$$q_1\int w_2(t)\mathbf{Q}_1(dt) < q_2\int w_1(t)\mathbf{Q}_2(dt).$$

Por consiguiente, en virtud del principio bayesiano, la regla bayesiana de decisión $\pi_Q(X)$ tendrá la forma $\pi_Q(X) = 1$ si

$$\int w_2(t)\mathbf{Q}_X(dt) < \int w_1(t)\mathbf{Q}_X(dt). \tag{1}$$

Supongamos que las distribuciones $\mathbf{Q}_i$ tienen densidades $q_i(t)$, $i = 1$, 2 respecto a la medida λ. Entonces, $\mathbf{Q}$ y la distribución a posteriori $\mathbf{Q}_x$ tendrán, respectivamente, densidades $q(t) = q_1q_1(t) + q_2q_2(t)$ y

$$q(t/X) = \frac{q(t)f_t(X)}{f(X)}, \quad f(X) = \int q(u)f_u(X)\lambda(du).$$

Esto significa que la relación (1) se puede escribir en la forma

$$q_1\int_{\Theta_1} w_2(t)q_1(t)f_t(X)\lambda(dt) < q_2\int_{\Theta_2} w_1(t)q_2(t)f_t(X)\lambda(dt). \tag{2}$$

El riesgo de la regla bayesiana $\pi_Q(X)$ es igual a

$$W(\pi_Q(\cdot), \theta) = w_1(\theta)\mathbf{M}_\theta\pi_Q(X) + w_2(\theta)(1 - \mathbf{M}_\theta\pi_Q(X)),$$

$$\bar{W}(\pi_Q(\cdot), \mathbf{Q}) = \int W(\pi_Q(\cdot), t)q(t)\lambda(dt).$$

Pasemos ahora a examinar las *alternativas semejantes*. Sea θ_1 cualquier valor fijo del parámetro θ. Al igual que en el § 3.14 supondremos que los conjuntos Θ_i tienen la forma siguiente:

$$\Theta_i = \theta_1 + \Gamma_i/\sqrt{n}, \tag{3}$$

donde Γ_i no depende de n. En lo que se refiere a Q_i, supondremos que éstas están inducidas por ciertas distribuciones Π_i concentradas en Γ_i y que no dependen de n. Si los conjuntos Γ_i están limitados, entonces, las estrategias de naturaleza θ estarán situadas en el $1/\sqrt{n}$-entorno del punto θ_1. Por eso, si $w_1(t)$, $w_2(t)$ son continuas y $w_i(t) > c > 0$, $i = 1, 2$ en los conjuntos Θ_2 y Θ_1, respectivamente, entonces, el juego estadístico $(\mathscr{D}, \Theta, W)$ para tal función de pérdidas no se distinguirá (según sus propiedades) del juego cuya función estadística de pérdidas constituye $w_i(t) = 1$ para $t \notin \Theta_i$ examinado en los §§ 3.14 y 3.15.

Aquí examinaremos una generalización más sustancial, análoga a la ejecutada en el § 6. Supondremos que la función de pérdidas $w(\delta_i, \theta) = w_i(\theta)$ depende de n de tal modo que

$$w_i(\theta) = w_{i,n}(\theta) = v_i(\sqrt{n}(\theta - \theta_1)), \tag{4}$$

donde $v_i(t)$ son funciones medibles limitadas que no dependen de n.

Siguiendo el § 3.14, llamaremos *problema* A al problema de búsqueda de la solución del juego $(\mathscr{D}, \Theta, W)$, descrito anteriormente, con ayuda de

la muestra $X \in \mathbf{P}$). Si se cumplen (3) y (4), hablaremos del problema A para hipótesis semejantes, con funciones de pérdidas $v_l(t)$.

Examinemos ahora otro juego estadístico $(\mathscr{D}_B, \Gamma, V)$ referente a la muestra $V \in \Phi_{\gamma, I^{-1}}$ de volumen unitario, donde $I = I(\theta_1)$ es la matriz de información de Fisher para la familia $\mathbf{P}_\theta$ en el punto θ_1. Este juego tiene el conjunto bipuntual de soluciones $D_B = (d_1, d_2)$ y el conjunto de estrategias de naturaleza (conjunto paramétrico) $\Gamma = \Gamma_1 \cup \Gamma_2$. La función de pérdidas $v(d, \gamma)$: $D_B \times \Gamma \to R$ se define por las relaciones

$$v(d_l, \gamma) = v_l(\gamma), \quad v_l(\gamma) = 0 \quad \text{para} \quad \gamma \in \Gamma_l.$$

Ahora bien, en este juego, $\mathscr{D}_B$ es la clase de todas las soluciones $d(Y)$: $\mathscr{Y} = R^k \to D_B$,

$$V(d(\cdot), \gamma) = v_1(\gamma)\Phi_{\gamma, I^{-1}}(d(Y) - d_1) + v_2(\gamma)\Phi_{\gamma, I^{-1}}(d(Y) = d_2)$$

(uno de los sumandos del segundo miembro es igual a cero). Análogamente se escriben las pérdidas para las estrategias randomizadas $\pi(Y)$ en los términos $\mathbf{M}\pi(Y)$, $Y \in \Phi_{\gamma, I^{-1}}$. Llamaremos *problema* B al problema enunciado.

Entre los problemas A y B aquí existe la misma relación que fue establecida entre estos problemas en el § 3.14. Sea $\pi(Y)$ la solución del problema B, óptima en uno u otro sentido (bayesiana o minimax). Y sea θ^* la e.v.m. en el problema A, $\gamma^* = (\theta^* - \theta_1)\sqrt{n}$. Entonces, $\pi(\gamma^*)$ será la solución asintóticamente óptima del problema A (en ese mismo sentido).

El "criterio límite de optimización" permite reducir el problema A a un problema más simple B.

Para que lo dicho adquiera sentido exacto daremos las definiciones siguientes. Supongamos que en Γ_l se dan las distribuciones Π_l. Pongamos $\Pi = q_1\Pi_1 + q_2\Pi_2$, $q_1 + q_2 = 1$ y designemos por $\mathbf{Q}$ la distribución en Θ, inducida por la distribución Π y por la transformación $\theta = \theta_1 + \gamma/\sqrt{n}$.

Definición 1. La solución $\pi_1(X)$ se llama *asintóticamente bayesiana* si

$$\lim_{n \to \infty} \sup\,[\tilde{W}(\pi_1(\cdot), \mathbf{Q}) = \tilde{W}(\pi_Q(\cdot), \mathbf{Q})] \leqslant 0.$$

Aquí, al igual que antes,

$$\tilde{W}(\pi(\cdot), \theta) = w_1(\theta)\mathbf{M}_\theta\pi(X) + w_2(\theta)(1 - \mathbf{M}_\theta\pi(X)),$$

$$\tilde{W}(\pi(\cdot), \theta) = \int \tilde{W}(\pi(\cdot), t)\mathbf{Q}(dt),$$

π_Q es la regla de decisión bayesiana.

Definición 2. La solución $\pi_1(X)$ se denomina *asintóticamente minimax* si para cualquier otra solución $\pi(X)$

$$\lim_{n \to \infty} [\sup_{\theta \in \Theta} \sup \tilde{W}(\pi_1(\cdot), \theta) - \sup_{\theta \in \Theta} \tilde{W}(\pi(\cdot), \theta)] \leqslant 0.$$

Aquí se podría comparar π_1 sólo con la regla minimax $\bar{\pi}$ (compárese con la definición 1).

Análogamente al § 3.14 también podríamos examinar las soluciones asintóticamente bayesianas y minimax en la clase $\mathbf{K}_\varepsilon$ de soluciones de las "pérdidas de primer género" asintóticas fijas:

$$\varepsilon = \lim_{n\to\infty} \sup \sup_{\theta\in\Theta_1} w_1(\theta)\mathbf{M}_\theta \pi(X).$$

Para obtener los resultados respectivos es suficiente comparar el contenido de este párrafo con el del § 3.14.

Designemos por $\pi_\Pi(Y)$ la solución bayesiana del juego $(\mathcal{D}_B, \Gamma, V)$ (o sea, del problema B), la cual corresponde a la distribución a priori Π, y supongamos, para abreviar, que los conjuntos Γ_i están limitados.

Teorema 1. *Supongamos que en el entorno del punto θ_1 se cumplen las condiciones (RR), y que las funciones v_i y la distribución Π_i son tales que* $0 < \int v_1(u)\Pi_2(du) < \infty$, $0 < \int v_2(u)\Pi_1(du) < \infty$. *Entonces, en las designaciones introducidas, el criterio*

$$\pi_1(X) = \pi_\Pi(\gamma^*), \quad \gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n}$$

será la solución asintóticamente bayesiana del juego $(\mathcal{D}, \Theta, W)$ (*o sea, del problema* A), *la cual corresponde a la distribución a priori* $\mathbf{Q}$.

Teorema 2. *Supongamos que en el entorno del punto θ_1 se cumplen las condiciones* (*RR*) *y que en el problema B existe la solución minimax $\bar{\pi}(Y)$ y la peor distribución correspondiente $\bar{\Pi}$. Entonces, el criterio* $\pi_1(X) = \bar{\pi}(\gamma^*)$ *será la solución asintóticamente minimax del problema* A.

Observación 1. Las condiciones del teorema de la existencia de $\bar{\pi}$ y $\bar{\Pi}$, en virtud de los teoremas del § 3, serán cumplidas siempre que v_i sean funciones continuas.

La **demostración** del teorema 1 es completamente análoga a la del teorema 3.14.1. De (2) se deduce que la regla bayesiana de decisión π_Q tendrá la forma $\pi_Q(X) = 1$ si

$$\frac{\int w_1(t)q_2(t)f_t(X)\lambda(dt)}{\int w_2(t)q_1(t)f_t(X)\lambda(dt)} > \frac{q_1}{q_2}. \tag{5}$$

Poniendo $Z_1(t) = \dfrac{f_{\theta_1+t}(X)}{f_{\theta_1}(X)}$ y teniendo en cuenta que

$$q_i(t)\lambda(dt) = \mathbf{Q}_i(dt), \quad \mathbf{Q}_i(\theta_1 + du/\sqrt{n}) = \Pi_i(du),$$
$$w_i(\theta_1 + u/\sqrt{n}) = v_i(u),$$

con ayuda de la sustitución de $t = \theta_1 + u/\sqrt{n}$ podemos transformar la desigualdad (5) reduciéndola a la forma

$$\frac{\int v_1(u)Z_1(u/\sqrt{n})\Pi_2(du)}{\int v_2(u)Z_1(u/\sqrt{n})\Pi_1(du)} = \frac{\int Z_1(u/\sqrt{n})\Pi_2'(du)}{\int Z_1(u/\sqrt{n})\Pi_1'(du)} > c, \quad c = \frac{q_1}{q_2}, \tag{6}$$

donde las distribuciones generalizadas $\Pi_l'(A) = \int_A v_{l+1}(u)\Pi_l(du)$ ($v_3(u) \equiv v_1(u)$, $l = 1, 2$) pueden ser transformadas, mediante renormalización, en probabilísticas, introduciendo las transformaciones $\Pi_l''(A) = \Pi_l'(A)/\Pi_l'(\Gamma_l)$ (según las condiciones $0 < \Pi_l'(\Gamma_l) < \infty$). Entonces, en calidad de (5) obtendremos la desigualdad que tiene exactamente la misma forma que en el § 3.14.

Los razonamientos ulteriores de la demostración se distinguen de los razonamientos respectivos del § 3.14 tan sólo por las simplificaciones. Esta tarea se la dejamos a cargo del lector. Dichos razonamientos se basan en la convergencia uniforme de ($\theta = \theta_1 + \gamma/\sqrt{n}$) en γ:

$$\tilde{W}(\pi_Q(\cdot), \theta) \to \tilde{V}(\pi_\Pi(\cdot), \gamma), \quad \tilde{W}(\pi_1(\cdot), \theta) \to \tilde{V}(\pi_\Pi(\cdot), \gamma), \tag{7}$$

donde $\pi_1(X) = \pi_\Pi(\gamma^*)$. ◁

Para demostrar el teorema 2 necesitaremos el

Lema 1. *Sea* Q *la distribución a priori, y* π_1, *la solución asintóticamente bayesiana que le corresponde, tal que*

$$\limsup_{n\to\infty} \tilde{W}(\pi_1(\cdot), Q) = c, \quad \limsup_{n\to\infty} \sup_{\theta\in\Theta} \tilde{W}(\pi_1(\cdot), \theta) \leqslant c. \tag{8}$$

Entonces, π_1 *es la solución asintóticamente minimax.*

Demostración. Al igual que antes, designemos por π_Q la solución bayesiana. Entonces, para cualquier solución π tenemos

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta} \tilde{W}(\pi, \theta) \geqslant \limsup_{n\to\infty} \tilde{W}(\pi, Q) \geqslant$$
$$\geqslant \limsup_{n\to\infty} \tilde{W}(\pi_Q, Q) \geqslant \limsup_{n\to\infty} \tilde{W}(\pi_1, Q) =$$
$$= c \geqslant \limsup_{n\to\infty} \sup_{\theta\in\Theta} \tilde{W}(\pi_1, \theta). \ ◁$$

Demostración del teorema 2. Sea $\bar{\Pi}$ la peor distribución en Γ, de modo que $\bar{\pi}(Y) = \pi_{\bar{\Pi}}(Y)$ sea la regla minimáx de decisión en el juego ($\mathscr{D}_B$, Γ, V). Entonces, según el teorema 1, $\pi_1(X) = \pi_{\bar{\Pi}}(\gamma^*)$ será la solución asintóticamente bayesiana para la distribución $\bar{Q}$ que corresponde a $\bar{\Pi}$, y para demostrar el teorema nos es suficiente convencernos que $\bar{Q}$ y π_1 satisfacen las condiciones del lema 1.

Designemos por $N_{\bar{\Pi}}$ el portador de la distribución $\bar{\Pi}$. Entonces, en vir-

tud de los teoremas del § 3.

$$\tilde{V}(\pi_{\bar{\Pi}}(\cdot),\ \gamma) = c \quad \text{para} \quad \gamma \in \mathrm{N}_{\Pi},$$
$$\sup_{\gamma\in\Gamma} \tilde{V}(\pi_{\bar{\Pi}}(\cdot),\ \gamma) \leqslant c. \qquad (9)$$

Pero para $\theta = \theta_1 + \gamma/\sqrt{n}$ tiene lugar (véase (7)) la convergencia $\tilde{W}(\pi_1(\cdot), \theta) \to \tilde{V}(\pi_{\bar{\Pi}}(\cdot),\ \gamma)$ uniforme en γ. De aquí y de (9) resulta (8). El teorema queda demostrado.

Suplemento I

Teoremas del tipo de Glivenko — Cantelli

En este Suplemento demostraremos las afirmaciones a base de las cuales se deducirán los teoremas 1.4.1. y 1.4.2. Utilizaremos, sin aclaraciones, las designaciones del párrafo 1.4 en el que estos teoremas han sido enunciados. Primero demostraremos la variante general auxiliar del teorema de Glivenko — Cantelli.

Definición 1. Llamaremos *aproximable finita* (respecto a la distribución P) la clase $\mathfrak{R}$ de conjuntos de $\mathfrak{B}_{\mathcal{X}} = \mathfrak{B}^m$, si cualquiera que sea $\varepsilon > 0$, para éste existe otra clase de conjuntos $\mathfrak{S}(\varepsilon)$, constituída por un número finito $N = N(\varepsilon)$ de elementos $S_1, \ldots, S_N$, $S_i \in \mathfrak{B}^m$, tal que para cualquier $B \in \mathfrak{R}$ habrá conjuntos A_1 y A_2 de $\mathfrak{S}(\varepsilon)$ dotados de las propiedades siguientes:

$$A_1 \subset B \subset A_2,$$
$$\mathbf{P}(A_2 - A_1) < \varepsilon. \tag{1}$$

Definamos sobre las clases de conjuntos, las operaciones de adición, de multiplicación y de complemento. Denominaremos clases $\mathfrak{R}_1 + \mathfrak{R}_2$ y $\mathfrak{R}_1\mathfrak{R}_2$ las clases de conjuntos del tipo $A \cup B$ y $A \cap B$, respectivamente, donde $A \in \mathfrak{R}_1$, $B \in \mathfrak{R}_2$. Llamaremos complemento $\overline{\mathfrak{R}}$ la clase de conjuntos formada por los complementos $\overline{A}$, $A \in \mathfrak{R}$.

Teorema 1. 1) *Supongamos que* $X_n = [X_\infty]_n$, $X_\infty \in \mathbf{P}$ *y que la clase* $\mathfrak{R}$ *es aproximable finita. Entonces*

$$\sup_{B \in \mathfrak{R}} |\mathbf{P}_n^*(B) - \mathbf{P}(B)| \underset{c.s}{\to} 0. \tag{2}$$

2) *La población de clases aproximables finitas está cerrada respecto a las operaciones introducidas.*

Demostración. La primera afirmación se obtiene con las mismas consideraciones que hemos usado en el caso unidimensional del teorema 1.22. Para los valores dados de $B \in \mathfrak{R}$ y $\varepsilon > 0$ existen $N = N(\varepsilon)$ y conjuntos A_1, A_2 dotados de la propiedad (1). Para ellos tenemos

$$\mathbf{P}_n^*(B) - \mathbf{P}(B) \leqslant \mathbf{P}_n^*(A_2) - \mathbf{P}(A_1) < \mathbf{P}_n^*(A_2) - \mathbf{P}(A_2) + \varepsilon,$$
$$\mathbf{P}_n^*(B) - \mathbf{P}(B) \geqslant \mathbf{P}_n^*(A_1) - \mathbf{P}(A_2) > \mathbf{P}_n^*(A_1) - \mathbf{P}(A_1) - \varepsilon.$$

Por eso

$$\bigcap_{k=1}^{N} \{ |\mathbf{P}_n^*(S_k) - \mathbf{P}(S_k)| < \varepsilon \} \subset \{ \sup_{B \in \mathfrak{R}} |\mathbf{P}_n^*(B) - \mathbf{P}(B)| < 2\varepsilon \},$$

donde $S_1, \ldots, S_N$ son los elementos de $\mathfrak{S}(\varepsilon)$. Como $\mathbf{P}_n^*(S_k) \underset{c.s}{\to} \mathbf{P}(S_k)$, de aquí ya sin dificultad obtenemos (2) (compárese con la demostración del teorema 1.2.2.A).

La segunda afirmación del teorema 3 es casi evidente. Supongamos que tenemos $\varepsilon > 0$ y que $\mathfrak{S}_1(\varepsilon_1)$ y $\mathfrak{S}_2(\varepsilon_2)$ son las clases aproximantes para $\mathfrak{R}_1$ y $\mathfrak{R}_2$, respectivamente. Sean, además, A y B conjuntos cualesquiera de $\mathfrak{R}_1$ y de $\mathfrak{R}_2$. De las relaciones $\varepsilon_1 + \varepsilon_2 = \varepsilon$,

$$A_1 \subset A \subset A_2, \quad \mathbf{P}(A_2 - A_1) < \varepsilon_1 \quad (A_i \in \mathfrak{S}_1(\varepsilon_1)),$$
$$B_1 \subset B \subset B_2, \quad \mathbf{P}(B_2 - B_1) < \varepsilon_2 \quad (B_i \in \mathfrak{S}_2(\varepsilon_2)),$$

obtenemos

$$A_1B_1 \subset AB \subset A_2B_2,$$
$$A_2B_2 - A_1B_1 \subset (A_2 - A_1) \cup (B_2 - B_1),$$
$$\mathbf{P}(A_2B_2 - A_1B_1) \leqslant \varepsilon.$$

Por lo tanto, la clase $\mathfrak{R}_1\mathfrak{R}_2$ es aproximable finita. La suma $\mathfrak{R}_1 + \mathfrak{R}_2$ y el complemento $\overline{\mathfrak{R}}$ se examinan análogamente. ◁

Corolario 1. *Sea* $\mathscr{X} = R^m$, $X_n = [X_\infty]_n \in F$. Entonces,

$$\sup_t |F_n^*(t) - F(t)| \xrightarrow{c.s.} 0$$

cuando $n \to \infty$, *donde* $F_n^*(t)$ *es la función empírica de distribución.*

Demostración. De la demostración del teorema 1.2.2A se deduce que las clases de subconjuntos $\mathfrak{R}_j = \{y \in R^m: y_j < t_j\}$, $-\infty < t_j < \infty$, para cada $j = 1, \ldots, m$, son clases aproximables finitas. En calidad del sistema $\mathfrak{S}(\varepsilon)$ es suficiente adoptar los semiespacios $\{y_j < z_k\}$ e $\{y_j \leqslant z_k\}$, $k = 1, \ldots, N$, donde z_k se han definido en (1.2.6).

Según la segunda afirmación del teorema 1, la clase de ángulos $\mathfrak{R} = \mathfrak{R}_1\mathfrak{R}_2 \ldots \mathfrak{R}_m$ también será aproximable finita. Nos queda hacer uso de la primera afirmación del teorema 1. ◁

El corolario 1 no es otra cosa sino el teorema 1.4.1..

Examinemos ahora las clases de conjuntos $\mathfrak{R}$ que satisfacen la condición siguiente (Γ). Sea K_M el cubo

$$K_M = \{y = (y_1, \ldots, y_m): \max_{1 \leqslant k \leqslant m} |y_k| \leqslant M\}.$$

(Γ) *Todos los conjuntos* $B \in \mathfrak{R}$ *poseen la siguiente propiedad: el* ε*-entorno* Γ_B^ε *de la frontera* $\Gamma_B = \partial(B \cap K_M)$ *tiene medida de Lebesgue (volumen)* $\mu(\Gamma_B^\varepsilon) \leqslant \varphi(\varepsilon, M)$, *donde* φ *sólo depende de sus argumentos, y para cualquier* M, $\varphi(\varepsilon, M) \to 0$ *cuando* $\varepsilon \to 0$.

Teorema 2. *Supongamos que* $\mathscr{X} = R^m$, $X \in \mathbf{P}$ *y la distribución* $\mathbf{P}$ *es absolutamente continua con respecto a la medida de Lebesgue. Entonces la clase* $\mathfrak{R}$ *que satisface la condición* (Γ) *es aproximable finita y, por consiguiente, para ella es válida* (2).

Demostración. Notemos antes que nada, que el problema cuyo espacio constituye R^m puede ser reducido al cubo K_M en el sentido siguiente. Supongamos que para cualquier M fija hay una clase $\mathfrak{S}$ de subconjuntos de K_M tal, que para cualquier $B' \in \mathfrak{R}$ y $B = B' \cap K_M$ se cumple (1). Entonces $\mathfrak{R}$ será aproximable finita. En efecto, para $\varepsilon > 0$, elegido en (1), hallemos $M = M(\varepsilon)$ tal, que $\mathbf{P}(K_M) \geqslant 1 - \varepsilon$, y pongamos $A_1' = A_1$, $A_2' = A_2 \cup \overline{K}_M$, donde A_i es un conjunto de (1), y $\overline{K}_M$ es el complemento hasta K_M. Entonces es evidente que

$$A_1' \subset B' \subset A_2', \quad \mathbf{P}(A_2' - A_1') \leqslant 2\varepsilon.$$

Así pues, podemos considerar que $\mathbf{P}(K_M) = 1$, $\mathfrak{R}$ consta de los sobconjuntos K_M.

Examinemos, en calidad de $\mathfrak{S}$, las figuras A_j formadas por distintas uniones de cubos cerrados, con aristas de longitud δ y con los vértices en los puntos

$$(j_1\delta, \ldots, j_m\delta), \quad -M/\delta < j_k < M/\delta, \quad k = 1, \ldots, m,$$

(para abreviar se puede admitir que δ divide totalmente (M). Definamos los conjuntos A_1,

A_2, respectivamente, como las uniones de todos los cubos que pertenecen y rozan con B. Es evidente que

$$A_1 \subset B \subset A_2,$$
$$\mu(A_2 - A_1) \leqslant \mu(\Gamma_B^{2\delta\sqrt{m}}) \leqslant \varphi(2\delta\sqrt{m}, M).$$

Eligiendo δ, el segundo miembro de esta desigualdad puede hacerse tan pequeño cuanto se quiera.

Seguidamente, $\mathbf{P}$ es en absoluto continua respecto a μ. Por eso, para ε dado se puede hallar $\gamma = \gamma(\varepsilon)$ tal, que $\sup_{\mu(A)<\gamma} \mathbf{P}(A) < \varepsilon$. Ahora, si δ se elige de tal modo que $\varphi(2\delta\sqrt{m}, M) < < \gamma$, entonces obtendremos

$$\mathbf{P}(A_2 - A_1) < \varepsilon. \ \triangleleft$$

Corolario 2. *La clase $\mathfrak{S}$ de todos los conjuntos convexos es aproximable finita y, por lo tanto, para* $\mathbf{P}$ *absolutamente continuas,*

$$\sup_{B\in\mathfrak{S}} |\mathbf{P}_n^*(B) - \mathbf{P}(B)| \xrightarrow[\text{c.s.}]{} 0.$$

En efecto, el "área" máxima de la superficie del conjunto convexo en K_M constituye $2m(2M)^{m-1}$ y equivale al "área" de la superficie K_M: y el volumen máximo $\mu((\partial K_M)^\varepsilon)$ del ε-entorno de ∂K_M no pasa de $2\varepsilon \cdot 2m(2M)^{m-1}$. Esto significa que se cumple la condición (Γ). $\triangleleft$

El corolario 2 coincide con el teorema 1.4.2. La observación en cuanto a la existencia de la condición de continuidad absoluta de $\mathbf{P}$ está presente en el § 1.4.

No es difícil notar que la condición (Γ) también será cumplida para las clases de conjuntos no convexos dotados de fronteras bastante suaves.

Suplemento II

Teorema funcional del límite para los procesos empíricos

Aquí demostraremos la afirmación siguiente (teorema 1.6.3). Sea

$$w^n(t) = \sqrt{n}(F_n^*(t) - t)$$

el proceso empírico definido en el § 1.6, y sea $w^0(t)$ el puente browniano.

Teorema 1. *Si f es una funcional medible: $D(0, 1) \to R$, continua en los puntos del espacio $C(0, 1)$ y en una métrica uniforme, entonces, cuando $n \to \infty$,*

$$f(w^n) \Rightarrow f(w^0).$$

Para demostrar el teorema necesitaremos dos lemas.

Lema 1. *Las distribuciones de dimensión finita de los procesos w^n convergen débilmente (cuando $n \to \infty$) hacia las distribuciones respectivas del proceso w^0.*

Demostración. Examinemos los vectores aleatorios de dimensión $(m + 1)$,

$$\mathbf{w}^n = (\Delta_0 w^n, \ldots, \Delta_m w^n),$$

donde, al igual que en el § 1.6, Δ_k designa las diferencias

$$\Delta_k g = g(t_{j+1}) - g(t_j),$$
$$t_{j+1} > t_j, \quad j = 0, \ldots, m, \quad t_0 = 0, \quad t_{m+1} = 1.$$

Designemos por $\mathbf{w}^0$ el vector análogo para el proceso $w^0(t)$. En virtud del segundo teorema de continuidad, para demostrar el lema es suficiente mostrar que $\mathbf{w}^n \Rightarrow \mathbf{w}^0$.

Hallemos las funciones características w^n y w^0. Para el vector $u = (u_0, \ldots, u_m)$ tenemos

$$\mathbf{M}e^{iw^0u^T} = \mathbf{M}\exp\left\{i\sum_{j=0}^{m} u_j\Delta_j w^0\right\} = \mathbf{M}\exp\left\{i\sum_{j=0}^{m} u_j(\Delta_j w - w(1)\Delta_j\right\},$$

donde $\Delta_j = t_{j+1} - t_j$, $j = 0, \ldots, m$, $w(t)$ es un proceso wieneriano estándar.

Representemos el exponente de la exponencial como una suma de magnitudes independientes. Para abreviar designemos $\sum\limits_{j=0}^{m} u_j\Delta_j = U$, obtendremos

$$\sum_{j=0}^{m} u_j(\Delta_j w - w(1)\Delta_j) = \sum_{j=0}^{m} (u_j - U)\Delta_j w.$$

En vista de que $\mathbf{M}e^{iuw(t)} = e^{-u^2t/2}$, entonces

$$\mathbf{M}e^{iw^0u^T} = \exp\left\{-\frac{1}{2}\sum_{j=0}^{m}(u_j - U)^2\Delta_j\right\} = \exp\left\{-\frac{1}{2}\left(\sum_{j=0}^{m} u_j^2\Delta_j - U^2\right)\right\}. \quad (1)$$

Ahora examinemos $\mathbf{M}e^{iw^nu^T}$. Sea, al igual que antes (véase el § 1.6),

$$\pi_n(t) = nF_n^*(t).$$

Entonces, como ya sabemos (véase (1.6.1)),

$$\mathbf{P}(\Delta_0\pi_n = k_0, \ldots, \Delta_m\pi_n = k_m) = \frac{n!}{k_0! \ldots k_m!}\Delta_0^{k_0} \ldots \Delta_m^{k_m}.$$

En el segundo miembro figuran los términos del desarrollo del polinomio $(\Delta_0 + \ldots + \Delta_m)^n$. Utilizando este argumento obtenemos

$$\mathbf{M}e^{i\sum_{j=0}^{m} u_j\Delta_j\pi_n} = \sum \frac{n!}{k_0! \ldots k_m}(e^{iu_0}\Delta_0)^{k_0} \ldots (e^{iu_m}\Delta_m)^{k_m} = (e^{iu_0}\Delta_0 + \ldots + e^{iu_m}\Delta_m)^n.$$

Como $\Delta_j w^n = \sqrt{n}(F_n^*(t_{j+1}) - F_n^*(t_j) - \Delta_j) = (\Delta_j\pi_n - n\Delta_j)/\sqrt{n}$, entonces

$$\mathbf{M}e^{iw^nu^T} = \exp\left\{-i\sum_{j=0}^{m} u_j\sqrt{n}\Delta_j\right\}\mathbf{M}\exp\left\{\frac{i}{\sqrt{n}}\sum_{j=0}^{m} u_j\Delta_j\pi_n\right\} = e^{i\sqrt{n}U}\left(\sum_{j=0}^{m} e^{iu_j/\sqrt{n}}\Delta_j\right)^n.$$

De aquí, para u fijo, utilizando las igualdades

$$e^\alpha = 1 + \alpha + \alpha^2/2 + O(\alpha^3), \quad \ln(1 + \alpha) = \alpha - \alpha^2/2 + O(\alpha^3),$$

cuando $\alpha = O(1)$, hallamos

$$\ln \mathbf{M}e^{iw^nu^T} = -iU\sqrt{n} + n\ln\left[1 - \sum_{j=0}^{m}(1 - e^{iu_j/\sqrt{n}})\Delta_j\right] =$$

$$= -iU\sqrt{n} + n\ln\left[1 + \sum_{j=0}^{m}\left(i\frac{u_j}{\sqrt{n}} - \frac{u_j^2}{2n} + O(n^{-3/2})\right)\Delta_j\right] =$$

$$= -iU\sqrt{n} + n\left[\frac{\sqrt{n}}{iU} - \frac{1}{2n}\sum_{j=0}^{m} u_j^2\Delta_j + \frac{U^2}{2n} + O(n^{-3/2})\right] =$$

$$= \frac{1}{2}\left[-\sum_{j=0}^{m} u_j^2\Delta_j + U^2\right] + O(n^{-1/2}).$$

Comparando con (1), vemos que cuando $n \to \infty$,

$$Me^{iw^n u^T} \to Me^{iw^0 u^T}. \tag{2}$$

Sólo queda utilizar el teorema de continuidad para las funciones características de las distribuciones multidimensionales (véase [11], p. 148). ◁

Lema 2. *Para cualquier* $\varepsilon > 0$

$$\lim_{n\to\infty} \sup \mathbf{P}(\omega_\Delta(w^n) > \varepsilon) \to 0 \tag{3}$$

para $\Delta \to 0$, *donde* $\omega_\Delta(y)$ *es el módulo de continuidad de la función* $y \in D(0, 1)$: $\omega_\Delta(y) = \sup_{\substack{0 \leqslant t_1 < t_2 \leqslant 1 \\ |t_1 - t_2| \leqslant \Delta}} |y(t_1) - y(t_2)|$.

Demostración. Sin limitar la generalidad, sólo podemos examinar los números binarios racionales $\Delta = 2^{-l}$. Para $m > 1$ tenemos

$$\omega_\Delta(w^n) \leqslant \omega_\Delta^{[m]} + 2 \max_{k \leqslant 2^m} \omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right),$$

donde

$$\omega_\Delta^{[m]} = \max_{\left|\frac{k-j}{2^m}\right| \leqslant \Delta} \left| w^n\left(\frac{k}{2^m}\right) - w^n\left(\frac{j}{2^m}\right)\right|,$$

$$\omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right) = \sup_{\frac{k-1}{2^m} \leqslant u < \frac{k}{2^m}} \left| w^n\left(\frac{k}{2^m}\right) - w^n(u)\right|.$$

Para demostrar (3) examinemos

$$\mathbf{P}(\omega_\Delta(w^n) \geqslant 3_\varepsilon) \leqslant \mathbf{P}\omega_\Delta^{[m]} \geqslant \varepsilon) + \mathbf{P}\left(\bigcup_{k=1}^{2^m}\left\{\omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right) \geqslant \varepsilon\right\}\right). \tag{4}$$

Aquí tomemos el primer sumando. Es fácil notar que cuando $l > 3$ el suceso

$$\bigcap_{r=l}^{m}\bigcap_{k=1}^{2^r}\left\{\left| w^n\left(\frac{k}{2^r}\right) - w^n\left(\frac{k-1}{2^r}\right)\right| < \frac{\varepsilon}{r^2}\right\}$$

provoca $\{\omega_\Delta^{[m]} < \varepsilon\}$. En vista de que para los sucesos adicionales tiene lugar la inclusión inversa, entonces

$$\mathbf{P}(\omega_\Delta^{[m]} \geqslant \varepsilon) \leqslant \mathbf{P}\left(\bigcup_{r=l}^{m}\bigcup_{k=1}^{2^r}\left\{\left| w^n\left(\frac{k}{2^r}\right) - w^n\left(\frac{k-1}{2^r}\right)\right| \geqslant \frac{\varepsilon}{r^2}\right\}\right). \tag{5}$$

Pero $w^n\left(\frac{k}{2^r}\right) - w^n\left(\frac{k-1}{2^r}\right) = \sqrt{n}\left(F_n^*\left(\frac{k}{2^r}\right) - F_n^*\left(\frac{k-1}{2^r}\right) - \frac{1}{2^r}\right)$, donde $n\left(F_n^*\left(\frac{k}{2^r}\right) - F_n^*\left(\frac{k-1}{2^r}\right)\right)$ es la frecuencia con que los elementos de la muestra van

a parar al intervalo cuya longitud constituye 2^{-r}. Con otras palabras, esta es la suma S_n de variables aleatorias en el esquema de Bernoulli con n pruebas y con una probabilidad del caso 1 igual a $p = 2^{-r}$. Como (véase [11], p. 105)

$$\mathbf{M}(S_n - np)^4 = n(p(1-p)^4 + (1-p)p^4) + 3n(n-1)p^2(1-p)^2 \leqslant np + 3n^2p^2,$$

entonces, según la desigualdad del tipo de Chébishev,

$$\mathbf{P}\left(\left|w^n\left(\frac{k}{2^r}\right) - w^n\left(\frac{k-1}{2^r}\right)\right| \geqslant \frac{\varepsilon}{r^2}\right) = \mathbf{P}\left(|S_n - np| \geqslant \frac{\varepsilon\sqrt{n}}{r^2}\right) \leqslant$$

$$\leqslant \frac{(np + 3n^2p^2)r^8}{\varepsilon^4 n^2} = \frac{r^8}{\varepsilon^4 2^r n} + \frac{3r^8}{\varepsilon^4 2^{2r}}.$$

Por consiguiente, el segundo miembro en (5) no supera

$$\sum_{r=1}^{m}\left[\frac{r^8}{\varepsilon^4 n} + \frac{3r^8}{\varepsilon^4 2^r}\right] \leqslant c\left(\frac{m^9}{\varepsilon^4 n} + \frac{l^8}{\varepsilon^4 2^l}\right),$$

donde c es cierta constante absoluta $\left(\sum_{r=1}^{m} r^8 \sim m^9/9\right.$ cuando $m \to \infty$ y $\left.\sum_{r=1}^{\infty} r^8 2^{-r} \sim 2l^8 2^{-l} \text{ cuando } l \to \infty\right)$. Poniendo $m = 3\log_2 n$, obtenemos que

$$\varlimsup_{n\to\infty} \mathbf{P}(\omega_\Delta^{[n]} \geqslant \varepsilon) \leqslant c\,\frac{l^8}{\varepsilon^4 2^l}.$$

Eligiendo l (o Δ), esta expresión puede hacerse tan pequeña como se quiera.

Ahora apreciemos el segundo sumando en (4), que no supera

$$2^m\mathbf{P}\left(\omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right) \geqslant \varepsilon\right). \tag{6}$$

El suceso que aquí figura bajo el signo de probabilidad significa que, elegido m, en el intervalo $((k-1)/n^3, k/n^3)$ cuya anchura es n^{-3}, la desviación de $n(F_n^*(u) - u)$ respecto a $n(F_n^*(k/n^3) - k/n^3)$ supera $\sqrt{n}\varepsilon$. En vista de que $\sqrt{n}\varepsilon \geqslant 3$, cuando n es bastante grande, para esto, en el intervalo $((k-1)/n^3, k/n^3)$ deben caer por lo menos 2 elementos de la muestra X, o sea, debe producirse el suceso $\{S_n \geqslant 2\}$ si volvemos a utilizar las designaciones para el esquema de Bernoulli cuando $p = n^{-3}$. Pero en vista de que $1 = (1 - p + p)^n = (1 - p)^n + np(1-p)^{n-1} + O(n^2p^2)$, entonces

$$\mathbf{P}(S_n \geqslant 2) = 1 - (1-p)^n - np(1-p)^{n-1} = O(n^2p^2).$$

Ahora bien, (6) no supera $n^3O(n^{-4}) = O(n^{-1}) = O(1)$. El lema queda demostrado. ◁

Demostración del teorema 1. Para cualquier $x \in D(0, 1)$ pongamos

$$\|x\| = \sup_{0\leqslant t\leqslant 1}|x(t)|,\ f_\varepsilon^+(x) = \sup_{\|y-x\|\leqslant\varepsilon} f(y),\ f_\varepsilon^-(x) = \inf_{\|y-x\|\leqslant\varepsilon} f(y)$$

y designemos por x_Δ la quebrada continua con nudos en los puntos $(k\Delta, x(k\Delta) = x_\Delta(k\Delta))$, $k = 0, \ldots, 1/\Delta$, donde Δ divide por completo 1. Es preciso señalar que

$$\|x - x_\Delta\| \leqslant \omega_\Delta(x) \tag{7}$$

y que $f_\varepsilon^\pm(x_\Delta)$ son funciones continuas del vector $(x(0), x(\Delta), x(2\Delta), \ldots, x(1))$. En virtud del

lema 1 y del segundo teorema de continuidad cuando $n \to \infty$,

$$f_\varepsilon^\pm(w_\Delta^n) \Rightarrow f_\varepsilon^\pm(w_\Delta^0). \tag{8}$$

Además, de la continuidad de w^0 y de la funcional f se deduce que

$$\|w_\Delta^0 - w^0\| \leq \omega_\Delta(w^0) \underset{P}{\to} 0 \quad \text{cuando} \quad \Delta \to 0, \tag{9}$$

$$f_\varepsilon^\pm(w^0) \underset{P}{\to} f(w^0) \quad \text{cuando} \quad \varepsilon \to 0. \tag{10}$$

De la definición de f_ε^- se desprende que $f_\varepsilon^-(y) \leq f(x)$ en el conjunto $\|y - x\| \leq \varepsilon$. Por eso

$$\mathbf{P}(f(w^n) \leq t) \leq \mathbf{P}(f_\varepsilon^-(w_\Delta^n) \leq t, \ \|w_\Delta^n - w^n\| \leq \varepsilon) + \mathbf{P}(\|w_\Delta^n - w^n\| > \varepsilon) \leq$$
$$\leq \mathbf{P}(f_\varepsilon^-(w_\Delta^n) \leq t) + \mathbf{P}(\omega_\Delta(w^n) > \varepsilon).$$

Pasando aquí al límite para $n \to \infty$ y utilizando (8) y (9), obtenemos

$$\limsup_{n\to\infty} \mathbf{P}(f(w^n) \leq t) \leq \mathbf{P}(f_\varepsilon^-(w_\Delta^0) \leq t) + \limsup_{n\to\infty} \mathbf{P}(\omega_\Delta(w^n) > \varepsilon). \tag{11}$$

Análogamente hallamos

$$\mathbf{P}(f_\varepsilon^-(w_\Delta^0) \leq t) \leq \mathbf{P}(f_{2\varepsilon}^-(w^0) \leq t) + \mathbf{P}(\omega_\Delta(w^0) > \varepsilon).$$

Sustituyamos ahora la última expresión en (11) y pasemos al límite cuando $\Delta \to 0$. Entonces, de (9) y del lema (2) obtenemos que

$$\limsup_{n\to\infty} \mathbf{P}(f(w^n) \leq t) \leq \mathbf{P}(f_{2\varepsilon}^-(w^0) \leq t).$$

De aquí y de (10) se deduce que

$$\limsup_{n\to\infty} \mathbf{P}(f(w^n) \leq t) \leq \mathbf{P}(f(w^0) \leq t).$$

Análogamente se establece la desigualdad inversa

$$\liminf_{n\to\infty} \mathbf{P}(f(w^n) < t) \geq \mathbf{P}(f(w^0) < t).$$

Las desigualdades obtenidas significan, evidentemente, que $f(w^n) \Rightarrow f(w^0)$. ◁

Examinemos otro teorema límite funcional para los procesos empíricos, el cual se asemeja mucho al teorema 1.

Supongamos que además de la muestra X de volumen n_1 tenemos una muestra Y de volumen n_2 que no depende de la primera y la cual procede de esa misma distribución uniforme en $[0, 1]$. En las condiciones de este apartado nos será más cómodo designar por $F_X^*(t)$ y $F_Y^*(t)$ las funciones empíricas de distribución de las muestras X e Y, respectivamente. Pongamos

$$w_{X,Y}(t) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,(F_X^*(t) - F_Y^*(t)).$$

Teorema 2. *Si la funcional f satisface las condiciones del teorema* 1, *entonces, para* $n_1 \to \infty$, *y* $n_2 \to \infty$

$$f(w_{X,Y}) \Rightarrow f(w^0).$$

Demostración. Demostremos este teorema utilizando la suposición simplificadora de que

$$a = \frac{n_2}{n_1 + n_2} \to a_0 \in [0, 1]$$

cuando $n \to \infty$. Tenemos

$$w_{X,Y}(t) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,[(F_X^*(t) - t) - (F_Y^*(t) - t)] = \sqrt{a}\,w_X(t) + \sqrt{1 - a}\,w_Y(t), \qquad (12)$$

donde $w_X(t)$ y $w_Y(t)$ son los procesos empíricos que corresponden a las muestras X e Y.

Como $\omega_\Delta(x + y) \leqslant \omega_\Delta(x) + \omega_\Delta(y)$, entonces, de (12) y del lema 2 se deduce inmediatamente el análogo del lema 2 para el proceso $w_{X,Y}(t)$: para cualquier $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup \mathbf{P}(\omega_\Delta(w_{X,Y}) > \varepsilon) \to 0.$$

La convergencia de las distribuciones de dimensión finita $w_{X,Y}$ y w^0 también se desprende de (12). En efecto, designemos por $w_{X,Y}$, w_X, w_Y los vectores construidos a base de los procesos $w_{X,Y}(t)$, $w_X(t)$, $w_Y(t)$, exactamente igual que como fue construido el vector w^n a base del proceso $w^n(t)$. Entonces, utilizando la independencia de X e Y y la demostración del lema 1, obtenemos

$$\mathbf{M}e^{i w_{X,Y} u^T} = \mathbf{M}e^{i\sqrt{a} w_X u^T}\mathbf{M}e^{i\sqrt{1-a} w_Y u^T} \to \exp\left\{\frac{a_0 + (1 - a_0)}{2} \times \right.$$

$$\left. \times \left(\sum_{j=0}^{m} u_j^2 \Delta_j - U^2\right)\right\} = \exp\left\{-\frac{1}{2}\left(\sum_{j=0}^{m} u_j^2\Delta_j - U^2\right)\right\} = \mathbf{M}e^{i w^0 u^T}.$$

En lo demás, la demostración del teorema 2 no se distingue en nada de la del teorema 1. ◁

Suplemento III

Propiedades de las esperanzas matemáticas condicionales

En el § 2.9 hemos citado las propiedades principales de las e.m.c. Más abajo aducimos las demostraciones de estas propiedades que siguen en el mismo orden que en el § 2.9.

1a. $\mathbf{M}(c\xi/\mathfrak{A}) = c\mathbf{M}(\xi/\mathfrak{A})$.

1b. $\mathbf{M}(\xi_1 + \xi_2/\mathfrak{A}) = \mathbf{M}(\xi_1/\mathfrak{A}) + \mathbf{M}(\xi_2/\mathfrak{A})$.

1c. *Si* $\xi_1 \leqslant \xi_2$ *c.s., entonces* $\mathbf{M}(\xi_1/\mathfrak{A}) \leqslant \mathbf{M}(\xi_2/\mathfrak{A})$ *c.s.*

Para demostrar la propiedad 1a es necesario convencerse, según la definición 2.9.2, de que

1) $c\mathbf{M}(\xi/\mathfrak{A})$ es una función $\mathfrak{A}$-medible.

2) $\mathbf{M}(c\mathbf{M}(\xi/\mathfrak{A}); A) = \mathbf{M}(c\xi; A)$ para cualquier $A \in \mathfrak{A}$.

El cumplimiento de la primera propiedad es evidente. La segunda propiedad se deduce de las propiedades de linealidad de una esperanza matemática ordinaria (o de una integral ordinaria):

$$\mathbf{M}(c\mathbf{M}(\xi/\mathfrak{A}); A) = c\mathbf{M}(\mathbf{M}(\xi/\mathfrak{A}); A) = c\mathbf{M}(\xi; A) = \mathbf{M}(c\xi; A).$$

La propiedad 1b se demuestra exactamente igual.

Para demostrar la propiedad 1c pongamos, para abreviar, $\tilde{\xi}_i = \mathbf{M}(\xi_i/\mathfrak{A})$. Entonces, para cualquier $A \in \mathfrak{A}$,

$$\int_A \tilde{\xi}_1 d\mathbf{P} = \mathbf{M}(\tilde{\xi}_1; A) = \mathbf{M}(\xi_1; A) \leqslant \mathbf{M}(\xi_2; A) = \int_A \tilde{\xi}_2 d\mathbf{P},$$

$$\int_A (\tilde{\xi}_2 - \tilde{\xi}_1) d\mathbf{P} \geqslant 0.$$

De aquí se deduce que $\tilde{\xi}_2 - \tilde{\xi}_1 \geqslant 0$ c.s.

2. **Desigualdad de Chébishev.** *Si* $\xi \geqslant 0$, $x > 0$, *entonces*

$$\mathbf{P}(\xi) \geqslant x/\mathfrak{A}) \leqslant \frac{\mathbf{M}(\xi/\mathfrak{A})}{x}.$$

Esta propiedad se desprende de lc, ya que $\mathbf{P}(\xi \geqslant x/\mathfrak{A}) = \mathbf{M}(I_{\{\xi \geqslant x\}}/\mathfrak{A})$, donde I_A es el indicador del suceso A, y es válida la desigualdad $I_{\{\xi \geqslant x\}} \leqslant \xi/x$.

3. *Si* $\mathfrak{A}$ *y* $\sigma(\xi)$ *son independientes, entonces* $\mathbf{M}(\xi/\mathfrak{A}) = \mathbf{M}\xi$. Como $\hat{\xi} = \mathbf{M}\xi$ es una función $\mathfrak{A}$-medible, sólo nos queda comprobar la segunda condición de definición 2.9.2: para cualquier $A \in \mathfrak{A}$

$$\mathbf{M}(\hat{\xi};\ A) = \mathbf{M}(\xi;\ A).$$

La validez de esta igualdad se deduce de la independencia de las variables aleatorias I_A y ξ y de las relaciones

$$\mathbf{M}(\xi;\ A) = \mathbf{M}(\xi I_A) = \mathbf{M}\xi \cdot \mathbf{M}I_A = \hat{\xi}\mathbf{P}(A) = \mathbf{M}(\hat{\xi};\ A).$$

4. **Teorema de convergencia monótona.** *Si* $0 \leqslant \xi_n \uparrow \xi$ *c.s., entonces* $\mathbf{M}(\xi_n/\mathfrak{A}) \uparrow \mathbf{M}(\xi/\mathfrak{A})$ *c.s.* En efecto, de $\xi_{n+1} \geqslant \xi_n$ c.s. resulta $\hat{\xi}_{n+1} \geqslant \hat{\xi}_n$ c.s., donde $\hat{\xi}_n = \mathbf{M}(\xi_n/\mathfrak{A})$. Por eso existe una $\mathfrak{A}$-medible $\hat{\xi}$ tal que $\hat{\xi}_n \uparrow \hat{\xi}$ c.s. En virtud del teorema ordinario de convergencia monótona, para cualquier $A \in \mathfrak{A}$,

$$\int_A \hat{\xi}_n d\mathbf{P} \to \int_A \hat{\xi} d\mathbf{P}, \quad \int_A \xi_n d\mathbf{P} \to \int_A \xi d\mathbf{P}.$$

En vista de que los primeros miembros de estas relaciones coinciden, también coinciden los segundos. Esto precisamente significa que $\hat{\xi} = \mathbf{M}(\xi/\mathfrak{A})$.

5. *Si* η *es real y* $\mathfrak{A}$ *es medible, entonces*

$$\mathbf{M}(\eta\xi/\mathfrak{A}) = \eta\mathbf{M}(\xi/\mathfrak{A}). \tag{1}$$

Si $\eta = I_B$ (indicador del conjunto $B \in \mathfrak{A}$), entonces, la afirmación es justa, ya que para cualquier $A \in \mathfrak{A}$

$$\int_A \mathbf{M}(I_B\xi/\mathfrak{A})d\mathbf{P} = \int_A I_B \xi d\mathbf{P} = \int_{AB} \xi d\mathbf{P} = \int_{AB} \mathbf{M}(\xi/\mathfrak{A})d\mathbf{P} = \int_A I_B \mathbf{M}(\xi/\mathfrak{A})d\mathbf{P}.$$

De aquí y de la linealidad de las e.m.c. resulta que la afirmación también es válida para cualesquiera funciones simples η.

Si $\xi \geqslant 0$ y $\eta \geqslant 0$, entonces, tomando la sucesión de funciones simples $0 \leqslant \eta_n \uparrow \eta$ y haciendo uso del teorema de convergencia monótona en la igualdad

$$\mathbf{M}(\eta_n\xi/\mathfrak{A}) = \eta_n\mathbf{M}(\xi/\mathfrak{A}),$$

obtenemos (1). El paso al caso de ξ y η arbitrarias se realiza ordinariamente: examinando las partes positivas y negativas de las variables aleatorias ξ y η. En este caso, para que las diferencias y sumas obtenidas tengan sentido, es necesario exigir la existencia de $\mathbf{M}|\xi| < \infty$, $\mathbf{M}|\xi_\eta| < \infty$.

6. *La desigualdad de Caucky — Buniakovski*

$$\mathbf{M}(\xi_1\xi_2/\mathfrak{A}) \geqslant [\mathbf{M}(\xi_1^2/\mathfrak{A})\mathbf{M}(\xi_2^2/\mathfrak{A})]^{1/2}$$

se demuestra exactamente igual que para las esperanzas matemáticas ordinarias (véase, por ejemplo, [11]), puesto que la demostración, además de la linealidad, no utiliza otras propiedades de las esperanzas matemáticas.

La desigualdad de Jensen

$$g(\mathbf{M}(\xi/\mathfrak{A})) \leqslant \mathbf{M}(g(\xi)/\mathfrak{A}) \qquad (2)$$

para cualquier función g convexa hacia abajo se deduce de las siguientes relaciones (compárese con [11]). En virtud de la convexidad de $g(x)$, para cada y habrá un número $g_1(y)$ tal, que

$$g(x) \leqslant g(y) + (x - y)g_1(y).$$

Pongamos aquí $x = \xi$, $y = \hat{\xi} = \mathbf{M}(\xi/\mathfrak{A})$ y tomemos la e.m.c. de ambos miembros de esta desigualdad. Como, en virtud de la propiedad 5,

$$\mathbf{M}[(\xi - \hat{\xi})g_1(\hat{\xi})/\mathfrak{A}] = g_1(\hat{\xi})\mathbf{M}[\xi - \hat{\xi}/\mathfrak{A}] = 0,$$

obtenemos (2).

7. *La fórmula de la probabilidad completa* se desprende de la propiedad 8 si en calidad de $\mathfrak{A}$ se adopta la σ-álgebra trivial.

8. *Si* $\mathfrak{A} \subset \mathfrak{A}_1 \mathfrak{F}$, entonces es válida la fórmula de "promediación sucesiva"

$$\mathbf{M}(\xi/\mathfrak{A}) = \mathbf{M}(\mathbf{M}(\xi/\mathfrak{A}_1)/\mathfrak{A}).$$

En efecto, para cualquier $A \in \mathfrak{A}$, en virtud de que $A \in \mathfrak{A}_1$,

$$\int_A \mathbf{M}(\mathbf{M}(\xi/\mathfrak{A}_1)/\mathfrak{A})d\mathbf{P} = \int_A \mathbf{M}(\xi/\mathfrak{A}_1)d\mathbf{P} = \int_A \xi d\mathbf{P} = \int_A \mathbf{M}(\xi/\mathfrak{A})d\mathbf{P}.$$

En conclusión cabe señalar que la propiedad 5 admite, para suposiciones amplias, la siguiente generalización.

5A. *Si* η *es medible respecto a* $\mathfrak{A}$, *y* $\varphi(\omega, \eta)$ *es la función medible de las variables* $\omega \in \Omega$ *y* $\eta \in R^k$, *entonces*

$$\mathbf{M}(\varphi(\omega, \eta)/\mathfrak{A}) = \psi(\omega, \eta), \text{ donde } \psi(\omega, y) = \mathbf{M}(\varphi(\omega, y)/\mathfrak{A}). \qquad (3)$$

Demostremos esta propiedad suponiendo que existe una sucesión de funciones simples η_n tal, que $\varphi(\omega, \eta_n)\uparrow\varphi(\omega, \eta)$, $\psi(\omega, \eta_n)\uparrow\varphi(\omega, \eta)$ c.s. En efecto, supongamos que $\eta_n = y_k$ para $\omega \in A_k \subset \mathfrak{A}$. Entonces

$$\varphi(\omega, \eta_n) = \sum_k \varphi(\omega, y_k)I_{A_k}.$$

En virtud de la propiedad 5, de aquí se deduce el cumplimiento de (3) para las funciones η_n. Queda utilizar el teorema de convergencia monótona (propiedad (4) en la igualdad

$$\mathbf{M}(\varphi(\omega, \eta_n)/\mathfrak{A}) = \psi(\omega, \eta_n).$$

Suplemento IV

Teorema de factorización de Neyman — Fisher

En este apartado demostraremos el teorema 2.12.1.

Para simplificar las designaciones supondremos, sin limitar la generalización, que $n = 1$ (pues la muestra X puede ser multidimensional). Además, en concordancia con el acuerdo de que el espacio probabilístico $(\mathscr{X}, \mathfrak{B})$ es muestral, escribiremos $\mathbf{P}_\theta(B)$ en vez de $\mathbf{P}_\theta(X \in B)$ y designaremos por l la dimensión de la estadística S.

Teorema 1. *Supongamos que se cumple la condición* (A_μ). *La estadística S es suficiente si y sólo si existe la función no negativa* $\psi(\theta, s)$ *medible respecto a* $s \in R^l$ *y la función no negati-*

va $h(x)$ medible respecto a $x \in \mathscr{X}$, tales que

$$f_\theta(x) = \frac{d\mathbf{P}_\theta}{d_\mu}(x) = \psi(\theta, S(x))h(x) \quad \text{c.d.} \quad [\mu]. \tag{1}$$

A la demostración del teorema 1 le antepondremos dos afirmaciones auxiliares. Introduzcamos en el planteamiento la

Condición (D). *La familia $\mathscr{P} = \{\mathbf{P}_\theta\}_{\theta \in \Theta}$ satisface la condición* (A_λ) (o sea, es dominada por la medida λ), *donde la medida probabilística λ tiene la forma siguiente*:

$$\lambda = \sum_i c_i \mathbf{P}_{\theta_i}, \theta_i \in \Theta, \quad c_i > 0, \quad \sum_i c_i = 1.$$

Teorema 2. *La condición (A_μ) es necesaria y suficiente para el cumplimiento de la condición D.*

Demostración. La necesidad es evidente. Demostremos la suficiencia. Sin limitar la generalidad se puede considerar que μ es una medida probabilística. En efecto, en vez de μ siempre se puede introducir la medida

$$\mu^*(A) = \sum_j \frac{\mu(AB_j)}{2^j_\mu(B_j)},$$

donde $\{B_j\}$ forma la partición del espacio $\mathscr{X}$, tal, que $\mu(B_j) < \infty$, $j = 1, 2, \ldots$

Sea $\bar{\mathscr{P}}$ la clase de todas las medidas probabilísticas de forma $\mathbf{P} = \sum c_i \mathbf{P}_{\theta_i}$, $\theta_i \in \Theta$, $c_i > 0$, $\sum c_i = 1$. Evidentemente, $\mathscr{P} \subset \bar{\mathscr{P}}$ y $\bar{\mathscr{P}}$ también satisface la condición (A_μ).

Designemos $p = d\mathbf{P}/d_\mu$ y examinemos la clase $\mathfrak{S}$ de conjuntos $C \in \mathfrak{B}$ para los cuales existe $\mathbf{P} \in \bar{\mathscr{P}}$ tal, que $p(x) > 0$ c.d. en C, $\mathbf{P}(C) > 0$. Sea $c_1, C_2, \ldots$ una sucesión de conjuntos de $\mathfrak{S}$ tal, que

$$\mu(C_i) \to \sup_{C \in \mathfrak{S}} \mu(C).$$

Como $C_i \in \mathfrak{S}$, entonces existe $\mathbf{P}^{(i)} \in \bar{\mathscr{P}}$ tal, que $p^{(i)} = \dfrac{d\mathbf{P}^{(i)}}{d_\mu} > 0$ c.s. en C_i. Pongamos

$$C_0 = \cup C_i, \quad \mathbf{P}^{(0)} = \sum_i c_i \mathbf{P}^{(i)}, \quad p^{(0)} = \sum_i c_i p^{(i)}$$

para cualesquiera $c_i > 0$, $\sum c_i = 1$. Es evidente que $p^{(0)} > 0$ en C_0 y, por lo tanto, $C_0 \in \mathfrak{S}$.

La afirmación del teorema quedará demostrada si determinamos que $\mathbf{P}^{(0)}(A) = 0$ contribuye a que $\mathbf{P}(A) = 0$ para todas $\mathbf{P} \in \bar{\mathscr{P}}$. Esto significará la continuidad absoluta de $\mathbf{P}_\theta$ respecto a $\lambda = \mathbf{P}^{(0)}$ y el cumplimiento de la condición (D).

Así pues, supongamos que $\mathbf{P}^{(0)}A = 0$ y que $\mathbf{P}$ es cualquier otro elemento de $\bar{\mathscr{P}}$. Designemos $C = \{x: p(x) > 0\}$. La afirmación requerida se deducirá de las tres relaciones siguientes:

$$\mathbf{P}(AC_0) = 0, \quad \mathbf{P}(A\overline{C_0}C) = 0, \quad \mathbf{P}(A\overline{C_0}) = 0,$$

donde $\overline{B}$ significa el complemento de B. La primera de estas relaciones se desprende del hecho de que $\mathbf{P}^{(0)}(AC_0) = 0$, $p^{(0)}(x) > 0$ en C_0 y, por lo tanto, $\mu(AC_0) = 0$. La segunda relación resulta del hecho de que $p(x) = 0$ en $\overline{C}$. Para demostrar la tercera relación admitamos que ella es injusta. Entonces, poniendo $R = A\overline{C_0}C$, obtenemos $\mu(R) > 0$, $\mu(C_0 \cup R) - \mu(C_0 > 0$. Pero esto contradice la igualdad

$$\mu(C_0) = \sup_{C \in \mathfrak{S}} \mu(C),$$

en vista de que $C_0 \in \mathfrak{S}$, $R \in \mathfrak{S}$, $C_0 \cup R \in \mathfrak{S}$. ◁

Ahora bien, hemos establecido que, al cumplirse las condiciones (A_a), existe una medida λ para la cual se cumple la condición (D).

Teorema 3. *La estadística S es suficiente si y sólo si existe una función medible $g_\theta(s)$ tal, que*

$$\frac{d\mathbf{P}_\theta}{d\lambda}(x) = g_\theta(S(x)) \text{ c.d. } [\lambda]. \tag{2}$$

Demostración. Para cualquier $B \subset R^l$ medible designemos $S^{-1}(B) = \{x \in \mathscr{X}: S(x) \in B\} \in \mathfrak{B}_{\mathscr{X}}$ y examinemos la distribución G_θ en R^l de la estadística S, inducida por la distribución $\mathbf{P}_\theta$:

$$G_\theta(B) = \int\limits_{S^{-1}(B)} \mathbf{P}_\theta(dx) = \int\limits_{S^{-1}(B)} \frac{d\mathbf{P}_\theta}{d\lambda}(x)\lambda(dx).$$

Examinemos también la distribución

$$\nu(B) = \int\limits_{S^{-1}(B)} \lambda(dx).$$

Por supuesto que G_θ es absolutamente continua respecto a ν, ya que $\nu(B) = 0$ contribuye a que $G_\theta(B) = 0$. Por eso existe una densidad $g_\theta(s)$ medible en s, tal, que

$$G_\theta(B) = \int\limits_B g_\theta(s)\nu(ds).$$

Ahora supongamos que S es una estadística suficiente y, por consiguiente, que existe una variante de distribución condicional $\mathbf{P}(A/s) = \mathbf{P}_\theta(A/S(x) = s)$ que no depende de θ. Según la definición de la distribución convencional, para cualquier $A_0 \in \sigma(S)$ se cumple

$$\int\limits_{A_0} \mathbf{P}(A/S(x))\mathbf{P}_\theta(dx) = \mathbf{P}_\theta(A \cap A_0).$$

De aquí también se deduce que

$$\int\limits_{A_0} \mathbf{P}(A/S)(x))\lambda(dx) = \lambda(A \cap A_0).$$

Esto significa que $\mathbf{P}(A/S)$ es a la vez una probabilidad condicional respecto a la distribución λ. Designemos esta probabilidad como e.m.c. $E_\lambda(I_A/S)$ del indicador I_A.

De (1), cuando $A_0 = R^l$, en virtud de las propiedades de la e.m.c., obtenemos

$$\mathbf{P}_\theta(A) = \int \mathbf{P}(A/S(x))\mathbf{P}_\theta(dx) = \mathbf{M}_\theta\mathbf{P}(A/S(X)) =$$
$$= \int \mathbf{P}(A/S)G_\theta(ds) = \int \mathbf{P}(A/S)g_\theta(s)\nu(ds) = \int \mathbf{P}(A/S(x))g_\theta(S(x))\lambda(dx) =$$
$$= \int E_\lambda(I_A/S(x))g_\theta(S(x))\lambda(dx) = \int E_\lambda(I_A g_\theta(S(x))/S(x))\lambda(dx) =$$
$$= \int I_A g_\theta(S(x))\lambda(dx) = \int\limits_A g_\theta(S(x))\lambda(dx).$$

Es evidente que esto significa precisamente (2).

Ahora supongamos que se cumple (2). Demostremos que la e.m.c. $E_\lambda(I_A/S)$, correspondiente a la distribución λ (que no depende de θ), es a la vez la e.m.c. $\mathbf{P}_\theta(A/S)$ para todas $\mathbf{P}_\theta \in \mathscr{P}$.

Para A y θ fijos introduzcamos la medida γ en $\mathfrak{B}$, definiéndola por la igualdad

$$\gamma(C) = \mathbf{P}_\theta(AC), \quad C \in \mathfrak{B},$$

así que $d\gamma/d\mathbf{P}_\theta = I_A$, $d\gamma/d\lambda = I_A g_\theta(S(x))$.

Para cualquier $C \in \sigma(S)$ tenemos

$$\gamma(C) = \int_C I_A \mathbf{P}_\theta(dx) = \mathbf{M}_\theta I_A I_C = \mathbf{M}_\theta I_C \mathbf{M}_\theta(I_A/S) = \int_C \mathbf{M}_\theta(I_A/S)\mathbf{P}_\theta(dx). \tag{3}$$

Por consiguiente, si γ, $\mathbf{P}_\theta$, λ se examinan como distribuciones en $\sigma(S)$, entonces

$$\frac{d\gamma}{d\mathbf{P}_\theta} = \mathbf{M}_\theta(I_A/S),$$

$$\frac{d\gamma}{d\lambda} = \mathbf{M}_\theta(I_A/S)\frac{d\mathbf{P}_\theta}{d\lambda} = \mathbf{M}_\theta(I_A/S)g_\theta(S).$$

Análogamente, en virtud de (3), en $\sigma(S)$,

$$\frac{d\gamma}{d\lambda} = \mathbf{E}_\lambda(I_A g_\theta(S)/S) = g_\theta(S)\mathbf{E}_\lambda(I_A/S).$$

De aquí se deduce que λ casi seguramente (aquí y más adelante, por λ y $\mathbf{P}_\theta$ entenderemos las distribuciones en $\sigma(S)$) constituirá

$$\mathbf{M}_\theta(I_A/S)g_\theta(S) = \mathbf{E}_\lambda(I_A/S)g_\theta(S). \tag{4}$$

Ahora hagamos uso de la propiedad (D), en virtud de la cual el cumplimiento de (4) λ c.s. significa el cumplimiento de esta relación por $\mathbf{P}_\theta$ c.s. Además, $\mathbf{P}_\theta$ c.s. es

$$g_\theta(S(x)) = \frac{d\mathbf{P}_\theta}{d\lambda}(x) \neq 0.$$

Por consiguiente, $\mathbf{P}_\theta$ c.s. es válida,

$$\mathbf{P}_\theta(A/S) = \mathbf{M}_\theta(I_A/S) = \mathbf{E}_\lambda(I_A/S).$$

Esto significa que la magnitud $\mathbf{E}_\lambda(I_A/S)$, que no depende de θ, puede ser elegida en calidad de probabilidad condicional $\mathbf{P}_\theta(A/S)$. ◁

Demostración del teorema 1. Si S es una estadística suficiente, entonces (1) se deduce del teorema 3, ya que

$$f_\theta(x) = \frac{d\mathbf{P}_\theta}{d\mu} = g_\theta(S(x))\frac{d\lambda}{d\mu}(x),$$

donde es preciso suponer que $g_\theta(s) = \psi(\theta, s)$, $\frac{d\lambda}{d\mu}(x) = h(x)$. Al contrario, si (1) es válida, entonces

$$\frac{d\lambda}{d\mu} = \sum c_i \frac{d\mathbf{P}_{\theta_i}}{d\mu} = \sum c_i \psi(\theta_i, S(x))h(x) = r(S(x))h(x).$$

Por eso, si $r(S(x)) > 0$, entonces

$$\frac{d\mathbf{P}_\theta}{d\lambda} = \frac{d\mathbf{P}_\theta}{d\mu}\cdot\frac{d\mu}{d\lambda} = \frac{\psi(\theta, S(x))}{r(S(x))}.$$

Si $r(S(x)) = 0$, entonces, $\frac{d\mathbf{P}}{d\lambda}(x)$ se puede definir arbitrariamente, ya que λ-medida y, por consiguiente, $\mathbf{P}_\theta$-medida del conjunto de tales puntos x es igual a cero. Poniendo $g_\theta(s) = \psi(\theta, s)/r(s)$ y aplicando el teorema 3 obtenemos que S es una estadística suficiente. ◁

Suplemento V

Ley de los grandes números y teorema central del límite. Variantes uniformes

1. Ley de los grandes números en el esquema de series. Examinemos las sucesiones $\{\xi_{k,n}\}_{k=1}^{n}$, $n = 1, 2, \ldots$, de vectores igualmente distribuidos en el esquema de series (la distribución $\xi_{k,n}$ depende de n) y supongamos que $\mathbf{M}\xi_{k,n} = 0$.

Designemos $\zeta_n = \sum_{k=1}^{n} \xi_{k,n}$.

Teorema 1. *Sea*

$$n\mathbf{M}|\xi_{k,n}| = a_n < a < \infty,$$
$$n\mathbf{M}(|\xi_{k,n}|;\ |\xi_{k,n}| > \tau) \to 0 \quad (1)$$

cuando $n \to \infty$ para cualquier $\tau > 0$. Entonces, para cualquier $\varepsilon > 0$,

$$\mathbf{P}(|\zeta_n| > \varepsilon) \to 0.$$

Demostración. Examinemos los cortes $\xi'_{k,n}$ de las variables aleatorias $\xi_{k,n}$ en el nivel τ:

$$\xi'_{k,n} = \begin{cases} \xi_{k,n}, & \text{si} \quad |\xi_{k,n}| \leqslant \tau, \\ 0, & \text{si} \quad |\xi_{k,n}| > \tau. \end{cases}$$

En virtud de la condición (1)

$$\mathbf{P}(\xi'_{1,n} \neq \xi_{1,n}) = \mathbf{P}(|\xi_{1,n}| > \tau) \leqslant \frac{1}{\tau}\mathbf{M}(|\xi_{1,n}|;\ |\xi_{1,n}| > \tau) = O(1/n),\ \mathbf{M}\xi'_{1,n} = O(1/n),$$

$$\mathbf{M}(\xi'_{1,n})^2 = \mathbf{M}(\xi^2_{1,n};\ |\xi_{1,n}| \leqslant \tau) \leqslant$$
$$\leqslant \tau\mathbf{M}(|\xi_{1,n}|;\ |\xi_{1,n}| \leqslant \tau) = \tau(a_n/n - \mathbf{M}(|\xi_{1,n}|;\ |\xi_{1,n}| > \tau)).$$

Por eso, para cualquier $\varepsilon > 0$ y para valores bastante grandes de n,

$$\mathbf{M}(\xi'_{1,n})^2 \leqslant 2a\tau/n,\ \mathbf{D}\xi'_{1,n} \leqslant 2a\tau/n,\ n\mathbf{M}\xi'_{1,n} < \varepsilon/2.$$

Pongamos $\zeta'_n = \sum_{j=1}^{n} \xi'_{j,n}$. Entonces, si los valores de n son bastante grandes,

$$\mathbf{P}(|\zeta_n| > \varepsilon) \leqslant \mathbf{P}\left(\bigcup_{j=1}^{n} \{\xi'_{j,n} \neq \xi_{j,n}\}\right) + \mathbf{P}(|\zeta'_n| > \varepsilon).$$

Aquí, el primer sumando no supera $n\mathbf{P}(\xi'_{1,n} \neq \xi_{1,n}) = o(1)$, y el segundo no pasa de

$$\mathbf{P}(|\zeta'_n - \mathbf{M}\zeta'_n| > \varepsilon/2) \leqslant 4\mathbf{D}\zeta'_n/\varepsilon^2 \leqslant 8a\tau/\varepsilon^2.$$

Como τ es arbitrario, para cualquier $\varepsilon > 0$ dado, el valor obtenido puede hacerse tan pequeño como se quiera. Eligiendo ahora un valor de n bastante grande, también podemos hacer tan pequeña como se quiera toda la probabilidad $\mathbf{P}(|\zeta_n| > \varepsilon)$. ◁

2. Teorema central del límite en el esquema de series. Aquí supondremos que

$$\mathbf{M}\xi_{j,n} = 0,\ \mathbf{M}|\xi_{j,n}|^2 < \infty.$$

Designemos $\sigma_n^2 = n\mathbf{M}\xi^T_{1,n}\xi_{1,n}$, $\zeta_n = \sum_{j=1}^{n} \xi_{j,n}$.

Teorema 2. *Supongamos que se cumplen las condiciones de Lindeberg*

$$n\mathbf{M}(|\xi_{1,n}|^2;\ |\xi_{1,n}| > \tau) \to 0$$

para $n \to \infty$ *para cualquier* $\tau > 0$. *Entonces, si* $\sigma_n^2 \to \sigma^2$,

$$\zeta_n \in \Phi_{0,\sigma^2}.$$

Corolario 1 (teorema central ordinario del límite). *Si* $\xi_1, \xi_2, \dots$ *es una sucesión de vectores independientes igualmente distribuidos*, $\mathbf{M}\xi_k = 0$, $\sigma^2 = \mathbf{M}\xi_k^T\xi_k < \infty$, $s_n = \sum_{k=1}^{n} \xi_k$, *entonces, para* $n \to \infty$,

$$\frac{s_n}{\sqrt{n}} \in \Phi_{0,\sigma^2}.$$

Esta afirmación es el corolario del teorema 2, ya que las variables aleatorias $\xi_{k,n} = \xi_k/\sqrt{n}$ satisfacen las condiciones del mismo.

Demostración del teorema 2. Examinemos las funciones características

$$\psi_n(t) = \mathbf{M}e^{i(t,\xi_{1,n})},\quad \varphi_n(t) = \mathbf{M}e^{i(t,\zeta_n)} = \psi_n^n(t).$$

Para demostrar el teorema necesitamos convencernos que para cualquier t

$$\varphi_n(t) \to \exp\left\{-\frac{1}{2}t\sigma^2t^T\right\}$$

cuando $n \to \infty$.

Hagamos uso de la variante unidimensional del teorema 1, demostrada en [11]. Las funciones $\psi_n(t)$ y $\varphi_n(t)$ pueden considerarse como funciones características

$$\psi_n^{(\omega)}(v) = \mathbf{M}e^{iv\xi_{1,n}^{\omega}} \quad \text{y} \quad \varphi_n^{(\omega)}(v) = \mathbf{M}e^{iv\zeta_n^{\omega}}$$

de las variables aleatorias $\xi_{1,n}^{\omega} = (\xi_{1,n}, \omega)$, $\zeta_n^{\omega} = (\zeta_n, \omega)$, donde $\omega = t/|t|$, $v = |t|$.

Mostremos que las variables aleatorias escalares $\xi_{k,n}^{\omega}$ satisfacen las condiciones del teorema 1 para el caso unidimensional. Es evidente que

$$\mathbf{M}\xi_{k,n}^{\omega} = 0,\ n\mathbf{M}(\xi_{1,n}^{\omega})^2 = n\mathbf{M}(\xi_{1,n}, \omega)^2 = \omega\sigma_n^2\omega^T \to \omega_{\sigma^2}\omega^T.$$

El cumplimiento de la condición de Lindeberg se deduce de la desigualdad evidente

$$n\mathbf{M}((\xi_{1,n}, \omega)^2;\ |(\xi_{1,n}, \omega)| > \tau) \leqslant n\mathbf{M}(|\xi_{1,n}|^2;\ |\xi_{1,n}| > \tau).$$

Ahora bien, para cualesquiera v y ω (o sea, para cualesquiera t)

$$\varphi_n(t)\mathbf{M}e^{iv\zeta_n^{\omega}} \to \exp\left\{-\frac{1}{2}v^2\omega\sigma^2\omega^T\right\} = \exp\left\{-\frac{1}{2}t\sigma^2t^T\right\}. \quad \triangleleft$$

3. Teoremas uniformes del límite para las sumas de las variables aleatorias que dependen del parámetro. En este apartado demostraremos los teoremas 29.1 y 29.2.

Sea $X \in \mathbf{P}_\theta$ y $a(x, \theta)$ una función medible $\mathscr{X} \times \Theta \to R^l$ dada,

$$s_n(\theta) = \sum_{j=1}^{n} a(x_j, \theta).$$

Diremos que la integral $a(\theta) = \int a(x, \theta)\mathbf{P}_\theta(dx)$ converge uniformemente en θ en la región $\Theta_0 \subset \Theta$ si

$$\sup_{\theta\in\Theta_0} \int_{|a(x,\theta)| > N} |a(x, \theta)|\mathbf{P}_\theta(dx) \to 0$$

cuando $N \to \infty$.

Teorema 3. (ley uniforme de los grandes números). *Si la integral* $a(\theta) = \int a(x, \theta)\mathbf{P}_\theta(dx)$ *converge uniformemente en* θ *en la región* $\Theta_0 \subset \Theta$, *entonces*

$$\zeta_n(\theta) = \frac{s_n(\theta)}{n} - a(\theta) \underset{\mathbf{P}}{\to} 0 \tag{2}$$

uniformementnte respecto a $\theta \in \Theta_0$.

Demostración. Supongamos que (2) no tiene lugar. Entonces habrá $\varepsilon > 0$, $\delta > 0$ y una sucesión $\theta_n \in \Theta_0$ tales, que

$$\mathbf{P}_{\theta_n}\left(\left|\frac{\zeta_n(\theta_n)}{n}\right| > \varepsilon\right) > \delta \tag{3}$$

para todos n.

Examinemos las variables aleatorias

$$\xi_{j,n} = \frac{n(x_j, \theta_n) - a(\theta_n)}{n}.$$

No es difícil notar que éstas satisfacen las condiciones del teorema 1. En efecto, pongamos $A_n = \{x;\ |a(x, \theta_n) - a(\theta_n)| > \tau n\}$. Entonces

$$n\mathbf{M}_{\theta_n}|\xi_{j,n}| \leqslant 2a = 2 \sup_{\theta \in \Theta_0} \int |a(x, \theta)|\mathbf{P}_\theta(dx) < \infty,$$

$$n\mathbf{M}_{\theta_n}(|\xi_{j,n}|;\ |\xi_{j,n}| > \tau) = \int_{A_n} |a(x, \theta_n) - a(\theta_n)|\mathbf{P}_{\theta_n}(dx) \to 0.$$

La última relación se deduce de la convergencia uniforme de la integral $a(\theta)$ y de la desigualdad de Chébishev

$$\mathbf{P}_{\theta_n}(A_n) \leqslant \frac{\mathbf{M}_{\theta_n}|\xi_{1,n}|}{\tau} \leqslant \frac{2a}{\tau n} \to 0.$$

Lo dicho significa que la sucesión $\xi_{j,n}$ satisface la ley de los grandes números:

$$\mathbf{P}_{\theta_n}\left(\left|\sum_{j=1}^{n} \xi_{j,n}\right| > \varepsilon\right) \to 0$$

para cualquier $\varepsilon > 0$. Esto contradice (3) y demuestra el teorema. ◁

Pasemos al *teorema central del límite*. Sea $\mathbf{M}_\theta a(x_1, \theta) = 0$.

Pongamos $\sigma^2(\theta) = \|\sigma_{ij}(\theta)\| = \mathbf{M}_\theta a^T(x_1, \theta)a(x_1, \theta)$ y designemos por $a_j(x, \theta)$, $j = 1, \ldots, l$ las coordenadas de los vectores $a(x, \theta)$.

Teorema 4 (Teorema central uniforme del límite). *Supongamos que las integrales* $\sigma_{jj}(\theta) = \mathbf{M}_\theta a_j^2(x_1, \theta)$ *convergen uniformemente en* $\Theta_0 \subset \Theta$, *o sea,*

$$\sup_{\Theta_0} \sigma_{jj}(\theta) < \infty$$

$$\sup_{\Theta_0} \mathbf{M}_\theta(a_j^2(x_1, \theta);\ |a_j(x_1, \theta)| > N) \to 0$$

cuando $N \to \infty$. *Entonces*

$$\frac{s_n(\theta)}{\sqrt{n}} \Subset \Phi_{0,\sigma^2(\theta)} \tag{5}$$

cuando $n \to \infty$ *uniformemente respecto a* $\theta \in \Theta_0$.

Demostración. El incumplimiento de (5) significará la existencia de una sucesión $\theta_n \in \Theta_0$ para la cual las sumas de las variables aleatorias $\xi_{j,n} = a(x_j, \theta_n)/\sqrt{n}$ no se aproximarán, según la distribución, a $\Phi_{0,\sigma^2(\theta_n)}$.

En virtud de la compactibilidad de la clausura $\{\sigma^2(\theta), \theta \in \Theta_0\}$, la sucesión θ_n puede considerarse elegida de tal modo que, para cierta matriz σ^2,

$$\sigma^2(\theta_n) = n\mathbf{M}_{\theta_n}\xi_{1,n}^T\xi_{1,n} \to \sigma^2. \tag{6}$$

Entonces, nuestra suposición acerca del incumplimiento de (5) significará que $\sum_{j=1}^{n} \xi_{j,n}$ no se aproximará, según la distribución, a Φ_{0,σ^2}. Pero esto es imposible en virtud del teorema 2, ya que $\xi_{j,n}$ satisfacen las condiciones del referido teorema. En efecto, en virtud de (6) es suficiente verificar la condición de Lindeberg. Para los conjuntos $A_{l,n} = \{|a_l(x_1, \theta_n)| > > \tau\sqrt{n}/l\}$ hallamos

$$\sup_{\theta\in\Theta_0} \mathbf{P}_\theta(A_{l,n}) \leqslant \sup_{\theta\in\Theta_0} \frac{l\sigma_{ll}(\theta)}{n\tau^2} \to 0$$

cuando $n \to \infty$. Utilizando el hecho de que $\{|\xi_{1,n}| > \tau\} \subset \bigcup_{l=1}^{l} A_{l,n}$, obtenemos

$$n\mathbf{M}_{\theta_n}(|\xi_{1,n}|^2;\ |\xi_{1,n}| > \tau) \leqslant \sum_{l,k=1}^{l} \mathbf{M}_{\theta_n}(a_l^2(x_1, \theta_n);\ A_{k,n}). \tag{7}$$

Aquí $\mathbf{M}_{\theta_n}(a_l^2(x_1, \theta_n); A_{l,n}) \to 0$ en virtud de la convergencia uniforme de la integral $\sigma_{ll}(\theta)$. Si $l \neq k$, entonces, poniendo $B_{l,N} = \{|a_l(x_1, \theta_n)| > N\}$, obtenemos

$$\mathbf{M}_{\theta_n}(a_l^2;\ A_{k,n}) = \mathbf{M}_{\theta_n}(a_l^2;\ A_{k,n}B_{l,N}) + \mathbf{M}_{\theta_n}(a_l^2;\ A_{k,n}\overline{B}_{l,N}).$$

Aquí, para $\varepsilon > 0$ dado se puede escoger N de tal modo que el primer sumando, en virtud de (4), sea menor que ε. El segundo sumando no supera $N^2\mathbf{P}_{\theta_n}(A_{k,n}) \to 0$ cuando $n \to \infty$. Esto significa que (7) converge a cero cuando $n \to \infty$.

Suplemento VI

Algunas afirmaciones referentes a las integrales que dependen del parámetro

1. Teoremas de la convergencia de las integrales que dependen del parámetro. Sea $\{\psi(t, y)\}$ una familia de funciones medibles que se dan en el espacio medible $(\mathscr{Y}, \mathfrak{B}_{\mathscr{Y}})$ con la medida ν en el. Nos interesarán las condiciones en las que

$$\int \psi(t, y)\nu(dy) \to \int \psi(\theta, y)\nu(dy) \quad \text{cuando} \quad t \to \theta. \tag{1}$$

Sea $\{A(t) = A(t, \theta), t \in \Theta\}$ cierta familia de conjuntos $\mathfrak{B}_{\mathscr{Y}}$. Designemos por $I_{A(t)}(x)$ el indicador $A(t)$, y por $\overline{A}(t)$, el complemento para $A(t)$.

La siguiente afirmación es cierta generalización del teorema conocido de Lebesgue.

Teorema 1. *Supongamos que la familia $\{A(t)\}$ es tal, que*

1) $\psi(t, y)I_{A(t)}(y) \to \psi(\theta)$, *cuando $t \to \theta$ para c.t.$[\nu]$ valores de y, para los cuales $\psi(\theta, y) \neq 0$.*

2) $\sup_t |\psi(t, y) I_{A(t)}(y)| \leqslant \psi(y)$, *donde* ψ *es la función integrable*

$$\int \psi(y)\nu(dy) < \infty.$$

Entonces, para que se cumpla (1) es necesario y suficiente que

$$\int \psi(t, y) I_{\overline{A(t)}}(y)\nu(dy) \to 0 \quad \text{cuando} \quad t \to 0. \tag{2}$$

Demostración. En virtud del teorema de Lebesgue,

$$\int \psi(t, y) I_{A(t)}(y)\nu(dy) \to \int \psi(\theta, y)\nu(dy).$$

En vista de que

$$\int \psi = \int \psi I_A + \int \psi I_{\overline{A}},$$

(1) es equivalente a (2). ◁

Si existe $\int \psi(\theta, y)\nu(dy)$, entonces, en calidad de conjunto $A(t)$ para c.t. $[\nu]$ de $\psi(t, y)$ continuas, se pueden utilizar los conjuntos

$$A(t) = \{y\colon |\psi(t, y)| \leqslant 2|\psi(\theta, y)|\},$$

así como se hace, por ejemplo, en la afirmación siguiente.

Corolario 1. *Sea* $\pi(x)$ *cualquier función medible limitada* $\mathscr{X}^n \to R$, $f_\theta(x)$, *continua en* θ *para c.t.* $[\mu^n]$ *valores de* $x \in \mathscr{X}^n$. *Entonces, la función*

$$\mathbf{M}_\theta \pi(X) = \int \pi(x) f_\theta(x) \mu^n(dx)$$

será continua en θ.

Demostración. Utilicemos el teorema 1 para $\mathscr{Y} = \mathscr{X}^n$, $y = x$, $\nu = \mu^n$, $\psi(t, x) = \pi(x) f_t(x)$, $A(t) = \{x\colon f_t(x) \leqslant 2 f_\theta(x)\}$. Es evidente que se han cumplido las condiciones 1) y 2). Como para $\pi(x) \equiv 1$, la función $\mathbf{M}_\theta \pi(X) \equiv 1$ es continua, entonces se cumple (véase (2))

$$\int_{x \notin A(t)} f_t(x) \mu^n(dx) \to 0$$

cuando $t \to \theta$. Pero de aquí, según el teorema 1, resulta la continuidad de $\mathbf{M}_\theta \pi(X)$ para cualquier función limitada π. ◁

Si sólo se trata de la condición suficiente para la convergencia (1) en caso de $\psi(t, y) \to \psi(\theta, y)$ c.d. cuando $t \to \theta$, en calidad de tal condición se puede utilizar la convergencia uniforme de las integrales en (1). Esta última puede ser definida como la existencia de una medida finita λ tal, que la desigualdad $\lambda(A) < \delta = \delta(\varepsilon)$ contribuye a que $\sup_t \int_A |\psi(t, y)| \nu(dy) < \varepsilon$ para $\varepsilon > 0$ dado.

Si existe la mayorante integrable $\psi(y) = \sup_t \psi(t, y)$, entonces siempre existe tal medida λ: es suficiente suponer que $\lambda(A) = \int_A \psi(y)\nu(dy)$.

2. Corolarios de las condiciones (*R*). Aquí demostraremos el lema 2.16.1 y la convergencia uniforme de la integral $I(\theta)$:

$$\sup_\theta \mathbf{M}_\theta(|l'(x_1, \theta)|^2;\ \ |l'(x_1, \theta)| > N) \to 0 \tag{3}$$

cuando $N \to \infty$ (precisamente tal uniformidad se tiene en cuenta en los §§ 2.24, 2.28 y 2.29). En vista de que los planteamientos referentes al parámetro unidimensional y multidimensional prácticamente no se distinguen, en este apartado y en el que le sigue nos limitaremos a estudiar el caso unidimensional.

Teorema 2. (lema 2.16.1). *Supongamos que se cumple la condición* (R) *y que* $S = S(X)$ *es cualquier estadística para la cual* $\mathbf{M}_\theta S^2 < c < \infty$ *cuando* $\theta \in \Theta$. *Entonces, en la igualdad*

$$a_S(\theta) = \mathbf{M}_\theta(S) = \int S(x) f_\theta(x) \mu^n(dx)$$

es posible la derivación bajo el signo integral:

$$a_S'(\theta) = \int S(x) f_\theta'(x) \mu^n(dx) = \mathbf{M}_\theta S L'(X, \theta), \tag{4}$$

siendo, en este caso, continua la función $a_S'(\theta)$.

Demostración. Nótese previamente que de (4), cuando $S(x) \equiv 1$ y $n = 1$, resulta

$$\int f_\theta'(x) \mu(dx) = 0. \tag{5}$$

Como $L'(X, \theta) = \sum_{i=1}^{n} l'(x_i, \theta)$ es la suma de las variables aleatorias independientes con media nula (véase (5)), entonces

$$\mathbf{D}_\theta L'(X, \theta) = \mathbf{M}_\theta(L'(X, \theta))^2 = n\mathbf{M}_\theta(l'(x_1, \theta))^2 = nI(\theta). \tag{6}$$

Ahora supongamos que la función

$$I_n(\theta) = \mathbf{M}_\theta(L'(X, \theta))^2 = 4\int (\sqrt{f_\theta(x)}')^2 \mu^n(dx)$$

es continua en θ (aún no podemos utilizar (6)). Hagamos uso ahora del teorema 1 para $\mathscr{Y} = \mathscr{X}^n$, $\nu = \mu^n$, $\psi(t, x) = (\sqrt{f_t(x)}')^2$, $\delta = t - \theta$, $A(t) = A_1(\delta) \equiv \{x: \sup_{v: |\theta - v| < |\delta|} \sqrt{f_v(x)} <$

$$< 2\sqrt{f_\theta(x)}, \quad \sup_{v: |\theta - v| < |\delta|} |\sqrt{f_v(x)}'| \leqslant 2|\sqrt{f_\theta(x)}'|\}. \tag{7}$$

Las condiciones 1) y 2) del teorema 1 para $\psi(x) = 2\psi(\theta, x)$ se cumplen en virtud de la continuidad de las funciones $\sqrt{f_\theta}$ y $\sqrt{f_\theta}'$. Por eso, de la convergencia de $I_n(t)$ hacia $I_n(\theta)$ cuando $t \to \theta$ obtenemos (véase (2)) que, cuando $t \to \theta$,

$$\varepsilon(t) = \int_{x \notin A_1(\delta)} (\sqrt{f_t(x)}')^2 \mu^n(dx) \to 0. \tag{8}$$

Al igual que como hemos obrado en el corolario 1, de aquí obtenemos la continuidad de $\int S(x) f_\theta'(x) \mu^n(dx)$. Para convencernos de ello es necesario valerse del teorema 1 "en sentido inverso" y utilizar los mismos conjuntos $A(t)$ y $\psi(t, x) = S(x) f_t'(x)$. Las condiciones 1) y 2) del teorema 1 serán, evidentemente, cumplidas ($\psi(x) = 2|S(x) f_\theta'(x)|$, $\int \psi(x) \mu^n(dx) \leqslant$ $\leqslant 4\mathbf{M}_\theta S^2 \times \int (\sqrt{f_\theta(x)}')^2 \mu^n(dx)$). El cumplimiento de (2) es asegurado por (8) y por la desigualdad recién citada, en la que la integración ha de efectuarse con arreglo al conjunto $x \notin A_1(\delta)$.

Ahora recurriremos directamente a la demostración de (4). Nótese que

$$\frac{1}{\delta}\left(\int S f_{\theta+\delta} \mu^n - \int S f_\theta \mu^n\right) = \int \int_0^1 S f_{\theta + u\delta}' \, du \, \mu^n = \int \int_0^1 2S\sqrt{f_{\theta+u\delta}}(\sqrt{f_{\theta+u\delta}})' \, du \mu^n.$$

Utilicemos de nuevo el teorema 1 para $\mathscr{Y} = R \times \mathscr{X}^n$, $y = (u, x)$, $\nu = \lambda \times \mu^n$ (λ es la medida de Lebesgue), $\psi(\delta, y) = S(x) f_{\theta+u\delta}'(x)$, $\delta \to 0$, $A(\delta) = A_1(\delta)$, donde $A_1(\delta)$ ha sido definida en (7). Otra vez de la continuidad de $\sqrt{f_\theta(x)}$ y $\sqrt{f_\theta(x)}'$ se deduce el cumplimiento de las condiciones 1) y 2) del teorema 1:

$$\psi(\delta, y) I_{A(\delta)}(x) \to S(x) f_\theta'(x) = \psi(0, y) \quad \text{cuando} \quad \delta \to 0,$$
$$\sup_\delta |\psi(\delta, y) I_{A(\delta)}(x)| \leqslant 4S(x)|f_\theta'(x)|,$$

donde, en virtud de la desigualdad de Cauchy — Buniakovski,

$$\int 4S|f_\theta'|\mu^n \leqslant 4\left[\int S^2 f_\theta \mu^n \cdot \int \frac{(f_\theta')^2}{f_\theta}\mu^n\right]^{1/2} < \infty.$$

Ahora bien, para demostrar (4) necesitaremos verificar la condición (2). Esta se desprende de la desigualdad de Cauchy — Buniakovski y de la relación (8):

$$\int\limits_{x \notin A_1(\delta)} \left|\int\limits_0^1 S\sqrt{f_{\theta+u\delta}}(\sqrt{f_{\theta+u\delta}})'\,du\mu^n\right| \leqslant$$

$$\leqslant \left[\int \int\limits_0^1 S^2 f_{\theta+u\delta}du\mu^n\right]^{1/2}\left[\int\limits_{x\notin A_1(\delta)}\int\limits_0^1 (\sqrt{f_{\theta+u\delta}})'^2 du\mu^n\right]^{1/2} \leqslant$$

$$\leqslant c^{1/2}\left[\int\limits_0^1 \varepsilon(\theta+u\delta)du\right]^{1/2} \to 0$$

cuando $\delta \to 0$.

Así pues, hemos demostrado (4) suponiendo que $I_n(\Theta)$ es continua. Pero para $n = 1$, $I_n(\theta) = I(\theta)$, esta suposición también se cumple en virtud de las condiciones (R). Por lo tanto, (4) es justa cuando $n = 1$ y, por consiguiente, también es justa (5). Pero de (5) resulta la relación (6) que significa la continuidad de $I(\theta)$. El teorema queda demostrado.

Teorema 3. *Si el conjunto Θ es compacto y la función $\sqrt{f_\theta(x)}$ para $[\mu]$ c.t. valores de x es continuamente derivable respecto a θ, entonces, la continuidad de $I(\theta)$ tendrá lugar si y sólo si se cumple* (3).

El teorema significa que la continuidad de $I(\theta)$ en la condición (R) puede ser sustituida por la condición (3).

Demostración. Supongamos que $f(\theta)$ se continua y que no se cumple (3). Entonces existe $\gamma > 0$, y las sucesiones $t \to \theta \in \Theta$ y $N_t \to \infty$ son tales que

$$m(t) \equiv \mathbf{M}_t[|l'(x_1, t)|^2;\ l'(x_1, t) > N_t] > \gamma \tag{9}$$

para todos los valores de t de la sucesión elegida.

Utilicemos el teorema 1 para $\mathscr{Y} = \mathscr{X}$, $\nu = \mu$, $\psi(t, x) = (\sqrt{f_t(x)}')^2 = \frac{1}{4}(l'(x, t))^2 f_t(x)$, $A(t) = \{x;\ |\sqrt{f_t(x)}'| \leqslant 2|\sqrt{f_\theta(x)}'|\}$. En virtud de la continuidad de $\sqrt{f_t(x)}'$, las condiciones 1) y 2) del teorema 1 se cumplen y, por consiguiente, de la continuidad de $I(t)$ se deducirá que

$$m_1(t) = \int\limits_{x\notin A(t)} |\sqrt{f_t(x)}'|^2\mu(dx) \to 0$$

cuando $t \to 0$. Pero $m(t) \leqslant m_1(t) + m_2(t)$, donde

$$m_2(t) = \int\limits_{B(t)\cap A(t)} (\sqrt{f_t}')^2\mu,\quad B(t) = \{x:\ 2|\sqrt{f_1(x)}'| > N_t\sqrt{f_1(x)}\}.$$

De la forma del conjunto $A(t)$ resulta

$$m_2(t) \leqslant 4\int\limits_{B(t)} |\sqrt{f_\theta}'|^2\mu.$$

Volviendo a utilizar la convergencia $(\sqrt{f_t(x)})' \to (\sqrt{f_\theta(x)})'$, $\sqrt{f_t(x)} \to \sqrt{f_\theta(x)}$ para $t \to \theta$, obtene-

mos que $B(t)$ converge hacia el conjunto de μ-medida 0. Esto significa que $\mu(B(t)) \to 0$, $m_2(t) \to 0$, $m(t) \to 0$ cuando $t \to \infty$. Hemos obtenido la contradicción con (9). La relación (3) queda demostrada.

Ahora supongamos que se cumple (3). En virtud del teorema 1, para demostrar la continuidad $I(t)$ es suficiente convencerse que con el mismo conjunto $A(t)$ que hemos utilizado más arriba, se cumple $m_1(t) \to 0$ cuando $t \to \infty$. Pero

$$m_1(t) \leqslant \int\limits_{|l'|>N} |l'|^2 f_t \mu + N^2 \int\limits_{x \notin A(t)} f_t \mu.$$

donde, por medio de la elección de N, la primera integral puede hacerse, en virtud de (3), tan pequeña como se quiera. Para estimar la segunda integral es necesario notar que $\mu(A(t)) \to 0$ y que cuando $C(t) = \{x; f_t(x) \leqslant 2f_\theta(x)\}$ se cumple $\int\limits_{x \notin C(t)} f_t \mu \to 0$ cuando $t \to 0$ (véase la demostración del corolario 1). Por eso

$$\int\limits_{x \notin A(t)} f_t \mu \leqslant 2 \int\limits_{x \notin A(t)} f_\theta \mu + \int\limits_{x \notin C(t)} f_t \mu \to 0$$

cuando $t \to 0$. ◁

3. **Corolarios de las condiciones (RR).**

Teorema 4. *Si se cumplen las condiciones* (RR), *entonces* $\int f_\theta''(x)\mu(dx) = 0$.

Junto con el teorema 2 esto asegura el cumplimiento de las condiciones (2.24.4) que necesitamos en el § 2.24.

Demostración. En virtud del teorema 2, para todos $\theta \in \Theta$,

$$\int f_\theta'(x)\mu(dx) = 0$$

y nos es suficiente demostrar que, cuando $t \to 0$,

$$J(t) = \frac{1}{t-\theta}\left[\int f_t'\mu - \int f_\theta'\mu\right] \to \int f_\theta''\mu.$$

Nótese que $\frac{1}{t-\theta}(f_t' - f_\theta') = \varphi_t f_t + \frac{f_\theta'}{f_\theta} \cdot \frac{f_t - f_\theta}{t-\theta}$, donde $\varphi_t = \frac{1}{t-\theta}\left(\frac{f_t'}{f_t} - \frac{f_\theta'}{f_\theta}\right)$. Aprovechando esta igualdad podemos representar $J(t)$ en forma de la suma de cuatro sumandos: $J(t) = J_1 + J_2 + J_3 + J_4$, donde

$$J_1 = \int \varphi_t f_\theta \mu, \quad J_2 = \int\limits_{l \leqslant N} \varphi_t (f_t - f_\theta)\mu,$$

$$J_3 = \int\limits_{l > N} \varphi_t (f_t - f_\theta)\mu, \quad J_4 = \int \frac{f_\theta'}{f_\theta} \cdot \frac{f_t - f_\theta}{t-\theta}\mu,$$

$l = l(x)$ es la mayorante para $l''(x, t)$ en las condiciones (RR). En virtud del teorema 2, cuando $n = 1$, $S(x) = l'(x, \theta)$ obtenemos

$$J_4 = \frac{1}{t-\theta}(\mathbf{M}_t l'(x_1, \theta) - \mathbf{M}_\theta l'(x_1, \theta)) \to \mathbf{M}_\theta (l'(x_1, \theta))^2 = I(\theta). \tag{10}$$

Seguidamente,

$$|\varphi_t| < l \tag{11}$$

y, por lo tanto, según el teorema de Lebesgue,

$$\lim_{t \to \theta} J_1 = \int \lim_{t \to \theta} \varphi_t f_\theta \mu = \int l'' f_\theta \mu = \int f_\theta'' \mu - I(\theta). \tag{12}$$

Volviendo a utilizar (11), obtenemos, en virtud de las condiciones (*RR*),

$$|J_3| \leqslant \int_{l \geqslant N} lf_{t\mu} + \int_{l \geqslant N} lf_{\theta\mu} \to 0$$

cuando $N \to \infty$. Por último, en virtud de la desigualdad de Cauchy — Buniakovski,

$$|J_2| \leqslant N\int |f_t - f_\theta|\mu \leqslant N\int \int_\theta^t |f_u'| du\mu \leqslant N \int_0^1 \sqrt{I(u)}du \to 0 \qquad (13)$$

cuando $t \to \theta$. Comparando (10)—(13) obtenemos que $0 = J(t) \to \int f_\theta^2\mu$. ◁

Suplemento VII

Desigualdades para la distribución de la relación de verosimilitud en el caso multidimensional

En este apartado demostraremos el siguiente teorema (teorema 28.2; las designaciones véanse en los §§ 2.21, 2.23, 2.28).

Teorema 1. *Supongamos que se cumplen las condiciones siguientes:*

$$\inf_u \frac{r(u)}{u^2} \geqslant g > 0, \qquad (1)$$

$$\mathbf{M}_\theta l'(x_1, \theta) = 0, \qquad (2)$$

$$\gamma = \sup_\theta \mathbf{M}_\theta |l'(x_1, \theta)|^s > \infty \qquad (3)$$

para cierto $s > k$. Entonces, para cualesquiera z, $n \geqslant 1$, $r > 0$,

$$\mathbf{P}_\theta\left(\sup_{|u| \geqslant r} Z\left(\frac{u}{\sqrt{n}}\right) \geqslant e^z\right) \leqslant c\gamma(e^{-z} + e^{-z/2})e^{-\beta g r^2},$$

donde $\beta > 0$ depende únicamente de k y s, $c < \infty$ depende de k, s y g.

Como ya hemos señalado en el § 2.28, para demostrar este teorema utilizaremos la posibilidad de estimar $\sup_{u \in K_{0,1}} p(u)$ para cierta función p y para el cubo unitario

$$K_{0,1} = \{u = (u_1, \ldots, u_k): 0 \leqslant u_j \leqslant 1, \ j = 1, \ldots, k\}$$

a través de los valores de $p(0)$ y $\int_{K_{0,1}} |p'(u)|^s du (p'(u) = \text{grad}\, p(u)$. Para realizar esta posibilidad necesitaremos la siguiente afirmación, cuya demostración reproducimos aquí, puesto que no figura en los conocidos manuales de análisis matemático. Por C_k, C_s y $C_{k,s}$ designaremos distintas constantes que sólo dependen de sus índices.

Lema 1. *Para cualquier $s > k$ existe $c_{k,s}$ tal, que*

$$\sup_{x \in K_{0,1}} |p(x)| \leqslant |p(x)| + c_{k,s}\left[\int_{K_{0,1}} |p'(x)|^s dx\right]^{1/s}$$

para cualquier $x \in K_{0,1}$.

Demostración. Para x, $y \in k_{0,1}$ es válida

$$p(x) = p(y) + \int_0^1 (p'(y + t(x - y)), x - y)dt.$$

Integrando esta igualdad respecto a $y \in K_{0,1}$, obtenemos

$$p(x) = \int_{K_{0,1}} p(y)dy + \int_{K_{0,1}} \int_0^1 (p'(y + t(x - y)),\ x - y)dtdy = I_1 + I_2, \tag{4}$$

donde I_1, I_2 designan el primero y el segundo, respectivamente. Sustituyamos en la integral I_2, las variables $y = \dfrac{z - tx}{1 - t}$. Entonces

$$y + t(x - y) = z, \quad x - y = \frac{x - z}{1 - t}, \quad dy = \frac{dz}{(1 - t)^k},$$

$$I_2 = \int_{K_{0,1}} (p'(z),\ x - z)K(x, z)dz, \tag{5}$$

donde $K(x, z) = \displaystyle\int_0^1 \varphi\left(\frac{z - tx}{1 - t}\right) \frac{dt}{(1 - t)^{k+1}}$, φ es el indicador del cubo $K_{0,1}$. Si aquí sustituimos $t = 1 - \dfrac{|z - x|}{u}$, entonces $\dfrac{1}{1 - t} = \dfrac{u}{|z - x|}$ y podemos escribir

$$K(x, z) = |z - x|^{-k} \int_{|z-x|}^{\infty} \varphi\left(x + \frac{z - x}{|z - x|}u\right) u^{k-1}du.$$

En vista de que para cualesquiera z, x, el portador de la función $\varphi\left(x + \dfrac{z - x}{|z - x|}u\right)$ está presente en el segmento $[0, 2\sqrt{k}]$, entonces

$$K(x, z) \leqslant |z - x|^{-k} \int_0^{2\sqrt{k}} u^{k-1}du = \frac{(2\sqrt{k})^k}{k|z - x|^k}.$$

Utilizando (5) y la desigualdad de Hölder, obtenemos

$$|I_2| \leqslant C_k \int_{K_{0,1}} \frac{|p'(z)|}{|z - x|^{k-1}} dz \leqslant c_k J \left(\int_{K_{0,1}} |p'(z)|^s dz\right)^{1/s},$$

donde

$$c_k = k^{-1}(2\sqrt{k})^k, \quad J = \left(\int_{K_{0,1}} \frac{dz}{|z - x|^{(k-1)r}}\right)^{1/r}, \quad \frac{1}{s} + \frac{1}{r} = 1.$$

Pero cuando $s > k$ se cumple $(k - 1)r = (k - 1)\dfrac{s}{s - 1} < k$,

$$J = J(k, s, x) < \left(\int_K \frac{dz}{|z|^{(k-1)r}}\right)^{1/r} = J(k, s) < \infty,$$

donde k es un cubo, o sea, $K = \{z\colon |z_j| \leqslant 1,\ j = 1, \ldots, k\}$.

Ahora bien, en virtud de (4),

$$\sup_{x \in K_{0,1}} |p(x)| \leqslant |I_1| + \sup_{x \in K_{0,1}} |I_2| \leqslant |p(x)| + 2 \sup_{x \in K_{0,1}} |I_2| \leqslant$$

$$\leqslant p(x) + 2c_k I(k, s) \left(\int_{K_{0,1}} |p'(z)|^s dz \right)^{1/s}.$$

El lema queda demostrado.

Así pues, la estimación de $\sup_{x \in K_{0,1}} |p(x)|$ es posible en los términos de $|p(x)|$ cuando está fijo $x \in K_{0,1}$ y $\int_{K_{0,1}} |p'(u)|^s du$ para $s > k$. Si seguimos el método que hemos utilizado en el caso unidimensional, ahora necesitaremos estimar $\mathbf{M}_\theta |p'(u)|^s$, donde, en calidad de $p(u)$ elegiremos la función

$$p(u) = Z^{1/s}(u). \tag{6}$$

Para esto, a su vez, necesitaremos los lemas siguientes.

Lema 2. *Sean* ξ_j, $j = 1, 2, \ldots,$ *los vectores independientes e igualmente distribuidos de* R^k, $\mathbf{M}\xi_1 = 0$, $\mathbf{M}|\xi_1|^s \leqslant \gamma < \infty$, $s \geqslant 2$. *Entonces*

$$\mathbf{M} \left| \sum_{j=1}^{n} \xi_j \right|^s \leqslant c_{k,s} \gamma n^{s/2}.$$

Demostración. Para simplificar los razonamientos nos limitaremos a examinar el caso cuando $s = 2m$ es un número entero par*. En este caso es suficiente examinar las variables aleatorias escalares ξ_j, puesto que $\xi_j = (\xi_{j,1}, \ldots, \xi_{j,k})$,

$$\left| \sum_{j=1}^{n} \xi_j \right|^{2m} = \left[\left(\sum_{j=1}^{n} \xi_{j,1} \right)^2 + \ldots + \left(\sum_{j=1}^{n} \xi_{j,k} \right)^2 \right]^m$$

y, en virtud de la desigualdad de Minkovski,

$$\left(\mathbf{M} \left| \sum_{j=1}^{n} \xi_j \right|^{2m} \right)^{1/m} \leqslant \left[\mathbf{M} \left(\sum_{j=1}^{n} \xi_{j,1} \right)^{2m} \right]^{1/m} + \ldots + \left[\mathbf{M} \left(\sum_{j=1}^{n} \xi_{j,k} \right)^{2m} \right]^{1/m}.$$

Para las ξ_j escalares tenemos

$$\mathbf{M} \left| \sum_{j=1}^{n} \xi_j \right|^s = \sum_{j_1, \ldots, j_n} \mathbf{M} \xi_1^{j_1} \ldots \xi_n^{j_n}, \tag{7}$$

donde la suma se realiza con arreglo a todos $j_1, \ldots, j_n$ enteros, tales, que $\sum_l j_l = s$, $j_l \neq 1$ ($j_l = 1$ se excluyen, ya que $\mathbf{M}\xi_j = 0$). Según la desigualdad de Hölder,

$$|\mathbf{M}\xi_l^{j_l}| \leqslant (\mathbf{M}|\xi_l|^s)^{j_l/s} = \gamma^{j_l/s}$$

y, por consiguiente,

$$\prod_{l=1}^{n} |\mathbf{M}\xi_l^{j_l}| \leqslant \prod_{l=1}^{n} \gamma^{j_l/s} = \gamma.$$

Nos queda estimar $\sum_{j_1, \ldots, j_n} 1$. Designemos por $(k_1, \ldots, k_p)$ los elementos no nulos ($k_l \geqslant 2$) del conjunto $(j_1, \ldots, j_n)$ $\left(\sum_{l=1}^{p} k_l = s \right)$. Entonces, la suma sujeta a estimación será igual a

---

*) La demostración en el caso general véase, por ejemplo, en [31], p. 255.

$\sum_{(k_1,\ldots,k_p)} A_p$, donde A_p es el número de ubicaciones de los elementos $k_1, \ldots, k_p$ en n lugares. Es evidente que $A_p \leqslant n(n-1)\ldots(n-p+1)$. El valor mayor posible de p es igual a $m = s/2$ (éste corresponde al conjunto $(2, 2, \ldots, 2)$), así que $A_p \leqslant A_m \leqslant n^m$. Pero el número de conjuntos diferentes $(k_1, \ldots, k_p)$ depende exclusivamente de s. Por consiguiente, la suma estimada no supera $c_s n^m$. ◁

Supongamos que la función $p(u)$ ha sido definida en (6).

Lema 3. *Si se cumplen las condiciones* (2) y (3),

$$\mathbf{M}_\theta |p'(u)|^s \leqslant c_s \gamma n^{s/2}.$$

Demostración

$$\mathbf{M}_\theta |p'(u)|^s = \mathbf{M}_\theta \left|\frac{1}{s} L'(X, \theta + u) Z^{1/s}(u)\right|^s =$$

$$= s^{-s}\mathbf{M}_\theta |L'(X, \theta + u)|^s Z(u) = s^{-s}\mathbf{M}_{\theta+u} |L'(X, \theta + u)|^s.$$

Nos queda utilizar el lema 2, aplicándolo a las variables aleatorias $\xi_j = l'(x_j, \theta + u)$.

Designemos por $K_{u,\Delta}$ el cubo en R^k, con lado de longitud Δ y con vértice en el punto $u = (u_1, \ldots, u_k)$:

$$K_{u,\Delta} = \{v \in R^k \colon u_i \leqslant v_i \leqslant u_i + \Delta,\ i = 1, \ldots, k\}.$$

Lema 4. *Si se cumplen las condiciones del teorema 1,*

$$\mathbf{P}_\theta\left(\sup_{v \in K_{u,\Delta}} Z\left(\frac{v}{\sqrt{n}}\right) > e^z\right) \leqslant c_{k,s}\gamma\Delta^k(e^{-z} + e^{-z/2})e^{-|u|^2 \beta g},$$

donde $\beta = \min\left(\frac{1}{4}, \frac{s-k}{4k}\right)$, $\Delta = e^{-|u|^2 g/4k}$.

Esta misma estimación será cierta para cualquier cubo con lado de longitud Δ *y que contiene el punto* u.

Demostración. Representemos el punto $v \in K_{u,\Delta}$ en forma de $v = u + t\Delta$, donde $t \in K_{0,1}$. Entonces

$$\mathbf{P} = \mathbf{P}_\theta\left(\sup_{v \in K_{u,\Delta}} Z\left(\frac{v}{\sqrt{n}}\right) > e^z\right) = \mathbf{P}_\theta\left(\sup_{t \in K_{0,1}} Z^{1/s}\left(\frac{u + t\Delta}{\sqrt{n}}\right) \geqslant e^{z/s}\right) =$$

$$= \mathbf{P}_\theta\left(\sup_{t \in K_{0,1}} p\left(\frac{u + t\Delta}{\sqrt{n}}\right) \geqslant e^{z/s}\right).$$

En virtud del lema 1,

$$\mathbf{P} \leqslant \mathbf{P}_\theta\left(p\left(\frac{u}{\sqrt{n}}\right) \geqslant \frac{1}{2}e^{z/s}\right) + \mathbf{P}_\theta\left(\left[\int_{K_{0,1}} \left|p_t'\left(\frac{u + t\Delta}{\sqrt{n}}\right)\right|^s dt\right]^{1/s} \geqslant\right.$$

$$\left.\geqslant \frac{1}{2c_{k,s}} e^{z/s}\right) = \mathbf{P}_{(1)} + \mathbf{P}_{(2)},$$

donde $\mathbf{P}_{(1)}$ y $\mathbf{P}_{(2)}$ designan el primero y el segundo sumandos, respectivamente. Estimemos $\mathbf{P}_{(1)}$ con ayuda de la desigualdad de Chébishev y del teorema 28.1:

$$\mathbf{P}_{(1)} \leqslant 2^{s/2}e^{-z/2}\mathbf{M}_\theta Z^{1/2}\left(\frac{u}{\sqrt{n}}\right) \leqslant 2^{s/2}e^{-z/2}e^{-\frac{|u|^2 g}{2}}. \tag{8}$$

Para estimar P_2 también utilizaremos la desigualdad de Chébishev:

$$P_{(2)} = P_\theta\left(\int_{K_{0,1}}\left(\frac{\Delta}{\sqrt{n}}\right)^s\left|p'\left(\frac{u+t\Delta}{\sqrt{n}}\right)\right|^s dt \geqslant \left[\frac{1}{2c_{k,s}}\right]^s e^z\right) \leqslant$$

$$\leqslant e^{-z}(2c_{k,s})^s \downarrow M_\theta \int_{K_{0,1}}\left(\frac{\Delta}{\sqrt{n}}\right)^s\left|p'\left(\frac{u+t\Delta}{\sqrt{n}}\right)\right|^s dt =$$

$$= (2c_{k,s})^s e^{-z}\left(\frac{\Delta}{\sqrt{n}}\right)^s \int_{K_{0,1}} M_\theta\left|p'\left(\frac{u+t\Delta}{\sqrt{n}}\right)\right|^s dt.$$

En virtud del lema 3,

$$P_{(2)} \leqslant c_{k,s}e^{-z}\left(\frac{\Delta}{\sqrt{n}}\right)^s \gamma n^{s/2} = c_{k,s}\gamma e^{-z}\Delta^s. \tag{9}$$

Poniendo

$$\Delta = e^{-\frac{|u|^2 g}{4k}}$$

y suponiendo, sin limitar la generalidad, $\gamma \geqslant 1$, obtenemos

$$P \leqslant 2^{s/2}e^{-z/2}e^{-\frac{|u|^2 g}{2}} + c_{k,s}\gamma e^{-z}\Delta^s \leqslant (2^{s/2} + c_{k,s})\gamma\Delta^k \times$$

$$\times\left[e^{-z/2}e^{-\frac{|u|^2 g}{4}} + e^{-z}e^{-\frac{|u|^2 g(s-k)}{4k}}\right] \leqslant (2^{s/2} + c_{k,s})\gamma\Delta^k(e^{-z/2} + e^{-z})e^{-\beta g|u|^2},$$

donde

$$\beta = \min\left(\frac{1}{4}, \frac{s-k}{4k}\right).$$

La última afirmación del lema se deduce, evidentemente, del lema 1 y de la demostración expuesta. El lema queda demostrado. ◁

Demostración del teorema 1. Cubramos todo el espacio R^k de un sistema de cubos $K_{u,\Delta}$ en los que las coordenadas de los puntos u son múltiplos de Δ. El número de tales cubos, que se intersecan con la capa $S_r = \{v \in R^k: r \leqslant |v| \leqslant r+1\}$, está limitado por la cantidad $c_k r^{k-1}$. Por lo tanto,

$$P_\theta\left(\sup_{v\in S_r} Z\left(\frac{v}{\sqrt{n}}\right) \geqslant e^z\right) \leqslant c_k r^{k-1}c_{k,s}\gamma(e^{-z/2} + e^{-z})e^{-r^2\beta g},$$

$$P_\theta\left(\sup_{|v|\geqslant r} Z\left(\frac{v}{\sqrt{n}}\right) \geqslant e^z\right) \leqslant c_k c_{k,s}\gamma(e^{-z/2} + e^{-z})\sum_{j=0}^{\infty}(r+j)^{k-1}e^{-(r+j)^2\beta g}.$$

La sucesión $(r+j)^{k-1}e^{-(r+j)^2\beta g}$ para todos $j \geqslant j(k, \beta g)$, donde $j(k, \beta g)$ depende únicamente de sus argumentos, decrece más rápidamente que la progresión geométrica con exponente $\frac{1}{2}$. Por eso, la serie en el segundo miembro de (10) no supera, para todos r, el primer sumando

con una exactitud de hasta la constante que sólo depende de k y βg. Como $\sup_{r\geq 0} r^{k-1} \times$
$\times\ e^{-\frac{1}{2} r^2 \beta g} < \infty$ también depende únicamente de k y βg, entonces

$$\mathbf{P}_\theta \left(\sup_{|v| \geq r} Z\left(\frac{v}{\sqrt{n}}\right) > e^z \right) \leq c\gamma(e^{-z/2} + e^{-z}) e^{-\frac{1}{2} r^2 \beta g},$$

donde c depende de k, s y βg. Sustituyendo aquí $\frac{\beta}{2}$ por β, obtenemos la afirmación del teorema. ◁

Suplemento VIII

Demostración de dos teoremas fundamentales de la teoría de los juegos estadísticos

Aquí vamos a suponer que se cumplen las condiciones siguientes.

Condición (A). *El conjunto de decisiones D y el conjunto de parámetros (estrategias puras de la naturaleza) Θ son espacios métricos compactos con métricas ϱ_D y ϱ_Θ, respectivamente.*

Condición (*B*). *La función de pérdidas $w(\delta, \theta)$: $D \times \Theta \to R$ es continua respecto a δ y θ en las métricas ϱ_D y ϱ_Θ, respectivamente.*

No necesitaremos la propiedad de $w(\delta, \theta) \geq 0$ y no supondremos que ésta tenga lugar.

Además, disponemos de la muestra $X \in \mathbf{P}_\theta$ de la distribución $\mathbf{P}_\theta$. Su volumen n, sin limitar la generalidad, se puede considerar igual a 1.

Condición (C). *Las distribuciones $\mathbf{P}_\theta$, con arreglo a la variación son continuas respecto a θ, o sea,*

$$\sup_{B \in \mathfrak{B}_{\mathscr{X}}} |\mathbf{P}_{\theta_m}(B) - \mathbf{P}_\theta(B)| \to 0$$

si $\varrho_\Theta(\theta_m, \theta) \to 0$ *cuando* $m \to \infty$.

Si se cumple la condición (A$_\mu$, o sea, si $\mathbf{P}_\theta$ tiene una densidad $f_\theta(x)$ respecto a cierta medida σ-finita μ en $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}})$:

$$f_\theta(x) = \frac{d\mathbf{P}_\theta}{d_\mu}(x),$$

entonces la condición (C) será equivalente a la continuidad de $f_\theta(x)$ en $L_1(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mu)$:

$$\int |f_{\theta_m}(x) - f_\theta(x)| \mu(dx) \to 0$$

si $\varrho_\Theta(\theta_m, \theta) \to 0$ cuando $m \to \infty$.

Las condiciones (A), (B) y (C) admiten, claro está, la posibilidad de ser finitas a los conjuntos D y Θ.

Si D es finito y consta de los puntos $\delta_1, \ldots, \delta_r$, entonces se cumplirá la condición A respecto a D (la elección de ϱ_D no tiene importancia), y la condición (B) significará la continuidad de las funciones $w(\delta_1, \theta), \ldots, w(\delta_r, \theta)$ respecto a ϱ_Θ.

Si ambos conjuntos D y Θ son finitos, las condiciones (A), (B) y (C) serán cumplidas automáticamente.

Designemos por σ_D y σ_Θ las σ-álgebras de los conjuntos de Borel de D y de Θ, respectivamente. Siguiendo el § 5.3, designemos por $(\tilde{\mathscr{D}}, \tilde{\Theta}, \tilde{W})$ el juego estadístico promediado, donde

como elementos de $\bar{\Theta}$ sirven las distribuciones Q en (Θ, σ_Θ), y como elementos de $\mathcal{D}$, las distribuciones $\pi(x) = \pi(x, \cdot)$ en (D, σ_D) (para cada $x \in \mathscr{X}$), donde $\pi(x, A)$ para cada $A \in \sigma_D$ es una función medible respecto a x.

La función de riesgo $\bar{W}(\pi, Q)$ es definida por la igualdad

$$\bar{W}(\pi, Q) = \int_{\Theta} \int_{\mathscr{X}} \int_{D} w(u, t)\pi(x, du)f_t(x)\mu(dx)Q(dt).$$

Si en vez del argumento Q se pone θ, entonces $\bar{W}(\pi, \theta)$ significará $\bar{W}(\pi, I_\theta)$, donde I_θ es la distribución concentrada en el punto θ. Este mismo acuerdo será válido respecto a la sustitución de $\pi \in \bar{\mathcal{D}}$, por $\delta \in \mathcal{D}$. También será más cómodo escribir W en vez de $\bar{W}$, ya que esto nunca conducirá a equivocaciones.

Lema 1. *Si se cumplen las condiciones* (A), (B), (C), *la función $W(\pi, \theta)$ será continua en θ para cualquier estrategia $\pi(x)$.*

Demostración. Tenemos para $\theta_n \to \theta$:

$$|W(\pi, \theta_n) - W(\pi, \theta)| \leqslant |M_\theta M[w(\pi(X), \theta) - w(\pi(X), \theta_n)/X]| +$$
$$+ |M_\theta M[w(\pi(X), \theta_n)/X] - M_{\theta_n} M[w(\pi(X), \theta_n)/X]| \leqslant$$
$$\leqslant \int |w(\pi(x), \theta) - w(\pi(x), \theta_n)| P_\theta(dx) + \sup_{\delta, \theta} |w(\delta, \theta)| \int |P_{\theta_n}(dx) - P_\theta(dx)|. \tag{1}$$

La primera integral aquí converge a 0 en virtud de la continuidad de la función w respecto a θ. La convergencia a cero de la segunda integral se deduce de la condición (C). En efecto, sea $f_{\theta_n}(x)$ la densidad P_{θ_n} respecto a la medida

$$\mu = P_\theta + \sum_{j=1}^{\infty} 2^{-j} P_{\theta_j},$$

y sea $B_n = \{x: f_{\theta_n}(x) \geqslant f_\theta(x)\}$. Entonces, la segunda integral en (1) será igual a

$$\int |f_{\theta_n}(x) - f_\theta(x)| \mu(dx) = 2 \int_{B_n} (f_{\theta_n}(x) - f_\theta(x))\mu(dx) = 2(P_{\theta_n}(B_n) - P_\theta(B_n)) \to 0.$$

El lema queda demostrado.

Teorema 1. (primer teorema fundamental). *Si se cumplen las condiciones* (A), (B) y (C), *el juego* ($\bar{\mathcal{D}}$, $\bar{\Theta}$, W) *tendrá precio y estrategias minimáx de ambos jugadores. Con otras palabras, existirá la distribución menos favorable $\bar{Q}$ y la regla minimáx de decisión $\bar{\pi}(x)$:*

$$W_* = \sup_Q \inf_\pi W(\pi, Q) = W(\bar{\pi}, \bar{Q}) = \inf_\pi \sup_Q W(\pi, Q) = W^*. \tag{2}$$

En virtud del lema 2.1, la afirmación (2) es equivalente al hecho de que

$$W(\bar{\pi}, \uparrow) = \sup_Q W(\bar{\pi}, \bar{Q}) = W(\bar{\pi}, \bar{Q}) = \inf_\pi W(\pi, \bar{Q}) = W(\downarrow, \bar{Q}). \tag{3}$$

Teorema 2 (segundo teorema fundamental). *Si se cumplen las condiciones* (A), (B) y (C), *las decisiones bayesianas $\pi_Q(x)$ formarán una clase completa. Con otras palabras, para cualquier $\pi_0 \in \bar{\mathcal{D}}$ habrá $Q \in \bar{\Theta}$, $\pi_Q \in \bar{\mathcal{D}}$ tales, que*

1) $W(\pi_Q, Q) = W(\downarrow, Q)$,

2) $W(\pi_Q, \theta) \leqslant W(\pi_0, \theta)$ *para todos* θ.

Demostración del teorema 2. El segundo teorema fundamental es el corolario del primero. Examinemos la estrategia arbitraria $\pi_0 \in \bar{\mathcal{D}}$ y el juego ($\bar{\mathcal{D}}$, $\bar{\Theta}$, W_0), donde W_0 se ha construido a base de la función $w_0(\delta, \theta) = w(\delta, \theta) - W(\pi_0, \theta)$, así que

$$W_0(\pi, \theta) = W(\pi, \theta) - W(\pi_0, \theta). \tag{4}$$

En virtud del lema 1, la función $v(\theta) = W(\pi_0, \theta)$ es continua en θ y, por lo tanto, la función de pérdidas $w_0(\delta, \theta) = w(\delta, \theta) - v(\theta)$, junto con $w(\delta, \theta)$, satisface la condición (B). Esto significa que el teorema 1 es aplicable al juego $(\mathscr{D}, \bar{\Theta}, W_0)$. En vista de que $W_0(\pi_0, \uparrow) = 0$ (véase (4)), el precio superior de este juego satisface la condición $W\S \leqslant 0$. Entonces, de (2) y (3) se deduce que existen $\bar{\pi}$, $\overline{Q}$ tales, que

$$\sup_{P} W_0(\bar{\pi}, P) = \sup_{\theta} W_0(\bar{\pi}, \theta) \leqslant 0, \; \bar{\pi} = \pi_{\overline{Q}}.$$

Estas dos relaciones son equivalentes a las afirmaciones 2) y 1) del teorema 2 si se pone $\overline{Q} = Q$, $\bar{\pi} = \pi_Q$. El teorema queda demostrado.

La demostración del teorema 1 se deducirá de los dos lemas siguientes.

Lema 2. *Al cumplirse las condiciones* (A), (B) y (C) *existirá una distribución* $\overline{Q}$ *tal, que* $W(\downarrow, \overline{Q}) \geqslant \inf W(\pi, \uparrow) = W^*$.

Lema 3. *Al cumplirse las condiciones* (A), (B) y (C) *existirá una estrategia* $\bar{\pi}$ *tal, que* $W(\bar{\pi}, \uparrow) \leqslant W^*$.

De las desigualdades de los lemas 2 y 3 se desprende la relación

$$W^* \geqslant W(\bar{\pi}, \uparrow) \geqslant W(\bar{\pi}, \overline{Q}) \geqslant W(\downarrow, \overline{Q}) \geqslant W^*$$

equivalente a (3) y, por consiguiente, a (2). Esto demuestra el teorema 1. ⊲

Los lemas 2 y 3 dividen la demostración del teorema 1 en dos partes. La primera de ellas (lema 2) está muy poco relacionada con el hecho de que el juego es estadístico. Esta parte de la demostración se realiza aproximadamente igual que para los juegos ordinarios (compárese con [31]).

Demostración del lema 2. Sea V un conjunto de funciones $\Theta \to R$ representables en forma de $v(\theta) = W(\pi, \theta)$, $\pi \in \mathscr{D}$. En virtud del lema 1, todas las funciones de V son continuas, así que $V \subset C(\Theta)$, donde $C(\Theta)$ es el espacio de todas las funciones continuas en Θ. Asimismo, sea $v_1(\theta) = W(\pi_1, \theta)$, $v_2(\theta) = W(\pi_2, \theta)$. En vista de que para $p \in [0, 1]$,

$$v(\theta) = pv_1(\theta) + (1 - p)v_2(\theta) = W(p\pi_1 + (1 - p)\pi_2, \theta),$$
$$\pi = p\pi_1 + (1 - p)\pi_2 \in \mathscr{D},$$

entonces, $v \in V$ y, por lo tanto, el conjunto V es convexo.

Ahora notemos que $W^* = \inf_{\pi} W(\pi, \uparrow) = \inf_{v \in V} \sup_{\theta} v(\theta)$. En vez de la función inicial $w(\delta, \theta)$ no será más cómodo examinar la función $\dfrac{w(\delta, \theta) - v_0 + 1}{W^* - v_0 + 1}$, $v_0 = \inf_{v \in V} \inf_{\theta} v(\theta)$. Designando la nueva función otra vez por $w(\delta, \theta)$ (en este caso el problema queda invariable), obtenemos que para ella

$$W^* = 1, \quad v_0 > 0. \tag{5}$$

Sea ahora U un conjunto de funciones continuas $v(\theta)$: $\Theta \to R$ tales, que $\sup_{\theta} v(\theta) < 1$. Es evidente que U es un conjunto abierto convexo de $C(\Theta)$. Además, de (5) se deduce que la intersección $V \cap U$ está vacía. Por eso, en virtud del teorema de Hahn — Banach (véase, por ejemplo, [31], p. 171, 200—206) existe una funcional lineal $L(v)$: $C(\theta) \to R$ tal, que

$$L(v) < 1 \quad \text{para} \quad v \in U, \quad L(v) \geqslant 1 \quad \text{para} \quad v \in V. \tag{6}$$

Esta funcional posee, cuando es necesario, la propiedad $L(v) \geqslant 0$ si $v(\downarrow) = \inf_{\theta} v(\theta) \geqslant 0$. En efecto, admitiendo la existencia del elemento $v_0 \in C(\Theta)$, $v_0(\downarrow) \geqslant 0$, para el cual $L(v_0) < 0$, obtenemos que $v_s = -sv_0 \in U$, cualquiera que sea $s > 0$, $L(v_s) = -sL(v_0) > 1$ y siempre que s sea bastante grande. Esto conduce a cierta contradicción con (6).

Pero la funcional no negativa L, en virtud del teorema de Riesz ([42)], p. 240), admite la representación en forma de la integral

$$L(v) = \int_{\Theta} v(\theta)\lambda(d\theta),$$

donde λ es una medida finita. Como $1 \geqslant \sup_{v \in U} L(v) = \lambda(\Theta)$, entonces, poniendo $\overline{Q}(A) = \lambda(A)/\lambda(\Theta)$, obtenemos para $v \in V$:

$$L(v) = \int W(\pi, \theta)\lambda(d\theta) = \lambda(\Theta) W(\pi, \overline{Q}),$$

$$W(\downarrow, \overline{Q}) = \frac{1}{\lambda(\Theta)} \inf_{v \in V} L(v) \geqslant 1 = W^*.$$

El lema queda demostrado.

Demostración del lema 3. En vista de que la función $W(\pi, \Theta)$ para cada $\pi \in \mathcal{G}$ es continua respecto a θ (véase el lema 1), nos es suficiente construir la estrategia $\bar{\pi}$ para la cual, con todos $k = 1, 2, \ldots,$

$$W(\bar{\pi}, \theta_k) \leqslant W^*, \tag{7}$$

donde θ_k son puntos de cierto conjunto numerable $T = \{\theta_1, \theta_2, \ldots\}$ siempre denso en D. Según la definición del precio superior de W^*, existe una sucesión de estrategias $\pi_n = \pi_n(x, \cdot)$ tal, que

$$W(\pi_n, \theta_k) < W^* + 1/n \tag{8}$$

para todos k.

Ahora, mediante las distribuciones π_n construyamos la sucesión de elementos aleatorios especialmente seleccionados ζ_n y separemos de ella la subsucesión convergente. Para esto, designemos por $f_{\theta_k}(x)$ la densidad de la distribución $\mathbf{P}_{\theta_k}$ respecto a la medida probabilística $\mu = \sum_{j=1}^{\infty} 2^{-k}\mathbf{P}_{\theta_k}$, así que

$$W(\pi_n, \theta_k) = \iint w(u, \theta_k)\pi_n(x, du) f_{\theta_k}(x)\mu(dx).$$

Examinemos el espacio $D \times R^T$, donde R^T es el espacio de los valores de los elementos $f(x) = \{f_{\theta_1}(x), f_{\theta_2}(x), \ldots\}$ con σ-álgebra $\mathfrak{B}^T$ engendrada por los conjuntos cilíndricos. Pongamos a cada estrategia π en correspondencia con el espacio probabilístico $(D \times \mathcal{X}, \sigma_D \mathfrak{B}_{\mathcal{X}}, \mathbf{P})$, donde la distribución $\mathbf{P}$ es definida por la igualdad

$$\mathbf{P}(\delta \in A, X \in B) = \int_B \mu(dx)\pi(x, A), \quad A \in \sigma_D, \quad A \in \mathfrak{B}_X. \tag{10}$$

Definamos en este espacio los elementos aleatorios $\zeta = \zeta(\delta; X) = (\delta; f_{\theta_1}(X), f_{\theta_2}(X), \ldots) = (\delta; f(X))$ y designamos por ζ_n los elementos correspondientes a π_O, así que ζ_n son variables aleatorias en el espacio probabilístico muestral $(D \times R^T, \sigma_D \times \mathfrak{B}^T, \Pi_n)$, y la distribución Π_n ha sido engendrada por π_n, por la fórmula (10) y por la aplicación $\zeta(\delta, x)$: $D \times \mathcal{X} \to D \times R^T$.

Designemos por $\Pi_n^{(k)}$ las contracciones de la distribución Π_n en $D \times R^k$ (es la distribución compatible $(\delta; f_{\theta_1}(X), \ldots, f_{\theta_k}(X))$, y por λ la distribución $f(X)$ en $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}}, \mu)$. Necesitaremos el

Lema 4. *Existe tal disitribución* $\overline{\Pi}$ *en el espacio medible* $(D \times R^T, \sigma_D \times \mathfrak{B}^T)$ *y tal subsucesión* $\{\pi_{n'}\} \subset \{\pi_n\}$, *que*

$$\Pi_{n'}^{(k)} \Rightarrow \overline{\Pi}^{(k)} \tag{11}$$

para cualquier k ($\overline{\Pi}^{(k)}$ *son las contracciones de* $\overline{\Pi}$),

$$\overline{\Pi}(D \times C) = \lambda(C), \quad C \in \mathfrak{B}^T. \tag{12}$$

La demostración del lema 4 se ofrecerá más tarde.

Designemos por $\bar{\zeta} = (\bar{\delta}; \bar{f})$ cierto elemento aleatorio con distribución $\overline{\Pi}$. La relación (12) significa que la distribución $\bar{f}$ coincide con λ (la segunda "coordenada" ζ_n no modifica la distribución al variar n). Como el espacio D constituye un compacto métrico, el mismo es separable y, por consiguiente, (véase [38], p. 191) existe cierta distribución condicional (regular) $\bar{\delta}$ respecto a $\bar{f}(X)$, la cual designaremos por $\Pi(\cdot/\bar{f}(x))$.

Examinemos la estrategia $\bar{\pi}(x, A) = \Pi(\bar{\delta} \in A/\bar{f}(X))$ y demostremos que para ella se cumple (7).

Señalemos previamente que

$$\mathbf{M} w(\bar{\delta}, \theta_k)\bar{f}_{\theta_k} = \mathbf{M}\bar{f}_{\theta_k}\mathbf{M}(w(\bar{\delta}, \theta_k)/X) = \int f_{\theta_k}(x)\int w(u, \theta_k)\bar{\pi}(u, dx)\mu(dx) = W(\bar{\pi}, \theta_k). \tag{13}$$

Seguidamente, en virtud del lema 4, la distribución $(\delta_{n^*}, f_{\theta_k}(X))$ converge débilmente hacia la distribución $(\bar{\delta}, \bar{f}_{\theta_k}(X))$. Como la función w es continua, la distribución compatible $(w(\delta_{n^*}, \theta_k), f_{\theta_k}(X))$ converge débilmente hacia la distribución $(w(\bar{\delta}, \theta_k), \bar{f}_{\theta_k}(X))$. Pero la función $g(u, v) = w(u, \theta_k)v$ es continua respecto a u y v y es mayorada por la función $g(v) = cv$, $c = \max_u w(u, \theta_k)$ tal, que $\mathbf{M}g(f_{\theta_k}(X)) = c\int f_{\theta_k}(x)\mu(dx) = c$. Por eso, según el teorema de continuidad para los momentos (véase el teorema 1.5.4),

$$\lim_{n^* \to \infty} \mathbf{M}g(\delta_{n^*}, f_{\theta_k}(X)) = \mathbf{M}g(\bar{\delta}, \bar{f}_{\theta_k}(X)),$$

o bien, que es lo mismo, $\lim\limits_{n^* \to \infty} \mathbf{M} w(\delta_{n^*}, \theta_k)f_{\theta_k}(X) = \mathbf{M} w(\bar{\delta}, \theta_k)\bar{f}_{\theta_k}(X)$.

En virtud de (9) y (13), esto nos ofrece la convergencia

$$\lim_{n^* \to \infty} W(\pi_n, \theta_k) = W(\bar{\pi}, \theta_k).$$

En vista de que el primer miembro de esta igualdad (véase (8)) no supera W^*, el lema 3 queda demostrado.

Demostración del lema 4. Fijemos cualquier $k \geqslant 1$ y examinemos $D \times R^k$ como espacio separable métrico completo respecto a la métrica engendrada por la métrica euclídea en R^k y la métrica ρ_D. Para cualquier $\varepsilon > 0$ en R^k habrá un compacto K_ε tal, que $\mathbf{P}((f_{\theta_1}(X), \ldots, f_{\theta_k}(X)) \in K_\varepsilon) \geqslant 1 - \varepsilon$. $D \times K_\varepsilon$ es un compacto un $D \times R^k$ y como

$$\mathbf{P}(\delta_n \in D, \quad (f_{\theta_1}(X), \ldots, f_{\theta_k}(X)) \in K_\varepsilon) \geqslant 1 - \varepsilon,$$

la sucesión de las distribuciones $\Pi_n^{(k)}$ es densa (véase [5]). Por consiguiente, según el teorema de Prójorov [5], existe una distribución $\overline{\Pi}^{(k)}$ y una subsucesión $n^{(k)} = (n_1^{(k)}, n_2^{(k)}, \ldots)$ tales, que $\Pi_{n^{(k)}}^{(k)} \Rightarrow \overline{\Pi}^{(k)}$. Pero las distribuciones $\overline{\Pi}^{(k)}$, evidentemente, se hallan en concordancia y, por consiguiente, según el teorema de Kolmogórov, en $(D \times R^T, \sigma_D \times \mathfrak{B}^T)$ existe cierta distribución $\overline{\Pi}$ para la cual $\overline{\Pi}^{(k)}$ son las contracciones en $(D \times R^k, \sigma_D \times \mathfrak{B}^k)$.

Por otro lado, podemos considerar que $n^{(k+1)} \subset n^{(k)}$. Poniendo $n^* = (n_1^{(1)}, n_2^{(2)}, n_3^{(3)}, \ldots)$ obtendremos una subsucesión para la cual $\Pi_{n^*}^{(k)} \Rightarrow \overline{\Pi}^{(k)}$ con todos los valores de k.

Demostremos ahora (12). Sea $C \in \mathfrak{B}^T$ un conjunto cilíndrico tal, que la $\overline{\Pi}$-medida de su frontera es igual a cero. Designemos por $C^{(k)} = C \cap R^k \in \mathfrak{B}^k$ el conjunto de R^k formado por las primeras k coordenadas de los puntos de C, y pongamos $\overline{C}^{(k)} = C^{(k)} \times R^{T-k} \in \mathfrak{B}^T$.

Entonces $\lambda(\overline{C}^{(k)}) = \Pi_{n^*}^{(k)}(D \times C^{(k)}) \to \overline{\Pi}^{(k)}(D \times C^{(k)})$. Como $\overline{C}^{(k+1)} \subset \overline{C}^{(k)}$, $C = \bigcap\limits_{k=1}^{\infty} \overline{C}^{(k)}$, entonces

$$\lambda(C) = \lim_{k \to \infty} \lambda(\overline{C}^{(k)}) = \lim_{k \to \infty} \overline{\Pi}^{(k)}(D \times C^{(k)}) = \lim_{k \to \infty} \overline{\Pi}(D \times \overline{C}^{(k)}) = \overline{\Pi}(D \times C).$$

El lema 4 queda demostrado.

Tabla I. Distribución normal $\Phi_{0,1}$

En la tabla se dan los valores de

$$\overline{\Phi}(x) = \Phi_{0,1}(x, \infty)) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt.$$

Tabla I

x	0	1	2	3	4
0,0	0,5000	0,4960	0,4920	0,4880	0,4840
0,1	,4602	,4562	,4522	,4483	,4443
0,2	,4207	,4168	,4129	,4090	,4052
0,3	,3821	,3783	,3745	,3707	,3669
0,4	,3446	,3409	,3372	,3336	,3300
0,5	,3085	,3050	,3015	,2981	,2946
0,6	,2743	,2709	,2676	,2643	,2611
0,7	,2420	,2389	,2358	,2327	,2297
0,8	,2119	,2090	,2061	,2033	,2005
0,9	,1841	,1814	,1788	,1762	,1736
1,0	,1587	,1562	,1539	,1515	,1492
1,1	,1357	,1335	,1314	,1292	,1271
1,2	,1151	,1131	,1112	,1093	,1075
1,3	,0968	,0951	,0934	,0918	,0901
1,4	,0808	,0793	,0778	,0764	,0749
1,5	,0668	,0655	,0643	,0630	,0618
1,6	,0548	,0537	,0526	,0516	,0505
1,7	,0446	,0436	,0427	,0418	,0409
1,8	,0359	,0351	,0344	,0336	,329
1,9	,0288	,0281	,0274	,0268	,0262
2,0	,0228	,0222	,0217	,0212	,0207
2,1	,0179	,0174	,0170	,0166	,0162
2,2	,0139	,0136	,0132	,0129	,0125
2,3	,0107	,0104	,0102	,0099	,0096
2,4	,0082	,0080	,0078	,0075	,0073
2,5	,0062	,0060	,0059	,0057	,0055
2,6	,0047	,0045	,0044	,0043	,0041
2,7	,0035	,0034	,0033	,0032	,0031
2,8	,0026	,0025	,0024	,0023	,0023
2,9	,0019	,0018	,0018	,0017	,0016
x =	3,0	3,1	3,2	3,3	3,4
$\overline{\Phi}(x)$ =	0,0013	0,0010	0,0007	0,0005	0,0003

Tabla I (continuación)

x	5	6	7	8	9
0,0	0,4810	0,4761	0,4721	0,4681	0,4641
0,1	,4404	,4364	,4325	,4286	,4247
0,2	,4013	,3974	,3936	,3897	,3859
0,3	,3632	,3594	,3557	,3520	,3483
0,4	,3264	,3228	,3192	,3156	,3121
0,5	,2912	,2877	,2843	,2810	,2776
0,6	,2578	,2546	,2514	,2483	,2451
0,7	,2266	,2236	,2206	,2177	,2148
0,8	,1977	,1949	,1922	,1894	,1867
0,9	,1711	,1685	,1660	,1635	,1611
1,0	,1469	,1446	,1423	,1401	,1379
1,1	,1251	,1230	,1210	,1190	,1170
1,2	,1056	,1038	,1020	,1003	,0985
1,3	,0885	,0869	,0853	,0838	,0823
1,4	,0735	,0721	,0708	,0694	,0681
1,5	,0606	,0594	,0582	,0571	,0559
1,6	,0495	,0485	,0475	,0465	,0455
1,7	,0401	,0392	,0384	,0375	,0367
1,8	,0322	,0314	,0307	,0301	,0294
1,9	,0256	,0250	,0244	,0239	,0233
2,0	,0202	,0197	,0192	,0188	,0183
2,1	,0158	,0154	,0150	,0146	,0143
2,2	,0122	,0119	,0116	,0113	,0110
2,3	,0094	,0091	,0089	,0087	,0084
2,4	,0071	,0069	,0068	,0066	,0064
2,5	,0054	,0052	,0051	,0049	,0048
2,6	,0040	,0039	,0038	,0037	,0036
2,7	,0030	,0029	,0028	,0027	,0026
2,8	,0022	,0021	,0021	,0020	,0019
2,9	,0016	,0015	,0015	,0014	,0014
$x =$	3,5	3,6	3,7	3,8	3,9
$\overline{\Phi}(x) =$	0,0002	0,0002	0,0001	0,0001	0,0000

Tabla II. Cuantilas de la distribución normal

En la tabla se dan los valores de λ_ε tales, que

$$\overline{\Phi}(\lambda_\varepsilon) = \Phi_{0,1}((\lambda_\varepsilon, \infty)) = \varepsilon.$$

Tabla II

100_ε	λ_ε	100_ε	λ_ε	100_ε	λ_ε
50	0,0000	20	0,8416	0,5	2,5758
45	0,1257	15	1,0364	0,1	3,0902
40	0,2533	10	1,2816	0,05	3,2905
35	0,3853	5	1,6449	0,01	3,7190
30	0,5244	2,5	1,9600	0,005	3,8906
25	0,6745	1	2,3263		

Tabla III. Distribución ji-cuadrado H_k

En la tabla se dan los valores (véanse el §§ 2.2)

$$H_k(x) = \mathbf{H}_k((x, \infty)) = \frac{1}{2^{k/2}\Gamma(k/2)} \int_x^\infty t^{k/2-1} e^{-t/2} dt$$

cuando $1 \leqslant k \leqslant 20$. Para mayores valores de k se puede utilizar la aproximación (véase el § 2.2, tabla I)

$$H_k(x) \approx \overline{\Phi}(\sqrt{2x} - \sqrt{2k-1}) \equiv \tilde{H}_k(x). \qquad (1)$$

La última columna de la tabla contiene los valores de $\tilde{H}_k(x)$ cuando $k = 20$. Comparándolos con los valores dados en la columna anterior se puede estimar el grado de precisión de la aproximación (1). Con el aumento de k disminuye el error.

Tabla III

x \ k	1	2	3	4	5
0,1	0,7518	0,9512	0,9918	0,9988	0,9998
,2	,6547	,9048	,9776	,9953	,9991
,4	,5271	,8187	,9402	,9825	,9953
,6	,4386	,7408	,8964	,9631	,9880
,8	,3711	,6703	,8495	,9385	,9770
1,0	,3173	,6065	,8013	,9098	,9626
,5	,2207	,4724	,6823	,8266	,9131
2	,1573	,3679	,5725	,7358	,8492
3	,0833	,2231	,3916	,5578	,7000
4	,0455	,1353	,2615	,4060	,5494
5	,0254	,0821	,1718	,2873	,4159
6	,0143	,0498	,1116	,1992	,3062
7	,0082	,0302	,0719	,1359	,2206
8	,0047	,0183	,0460	,0916	,1562
9	,0027	,0111	,0293	,0611	,1091
10	,0016	,0067	,0186	,0404	,0752
11	,0009	,0041	,0117	,0266	,0514
12	,0005	,0025	,0074	,0174	,0348
13	,0003	,0015	,0046	,0113	,0234
14	,0002	,0009	,0029	,0073	,0156
15	,0001	,0006	,0018	,0047	,0104
16	,0001	,0003	,0011	,0030	,0068
17		,0002	,0007	,0019	,0045
18		,0001	,0004	,0012	,0030
19		,0001	,0003	,0008	,0019
20		,0001	,0002	,0001	,0013
21			,0001	,0003	,0008
22			,0001	,0002	,0005
23				,0001	,0003
24				,0001	,0002
25				,0001	,0001

Tabla III (continuación)

x \ k	6	7	8	9	10
0,5	0,9978	0,9995	0,9999	1,0000	1,0000
1,0	,9856	,9948	,9983	0,9994	0,9998
,5	,9595	,9823	,9927	,9972	,9989
2,0	,9197	,9598	,9810	,9915	,9963
,5	,8685	,9271	,9617	,9809	,9909
3	,8089	,8850	,9344	,9643	,9814
4	,6767	,7798	,8571	,9114	,9474
5	,5438	,6600	,7576	,8343	,8912
6	,4232	,5398	,6472	,7399	,8153
7	,3204	,4284	,5366	,6371	,7254
8	,2381	,3326	,4335	,5342	,6288
9	,1736	,2527	,3423	,4373	,5321
10	,1246	,1886	,2650	,3505	,4405
11	,0884	,1386	,2017	,2757	,3575
12	,0620	,1006	,1512	,2133	,2851
13	,0430	,0721	,1119	,1626	,2237
14	,0296	,0512	,0818	,1223	,1730
15	,0203	,0360	,0592	,0909	,1321
16	,0138	,0251	,0424	,0669	,0996
17	,0093	,0174	,0301	,0487	,0744
18	,0062	,0120	,0212	,0352	,0550
19	,0042	,0082	,0149	,0252	,0403
20	,0028	,0056	,0103	,0179	,0193
21	,0018	,0038	,0072	,0127	,0211
22	,0012	,0025	,0049	,0084	,0151
23	,0008	,0017	,0034	,0062	,0108
24	,0005	,0011	,0023	,0043	,0076
25	,0003	,0008	,0016	,0030	,0054
26	,0002	,0005	,0011	,0020	,0037
27	,0002	,0003	,0007	,0014	,0026
28	,0001	,0002	,0004	,0010	,0018
29	,0001	,0002	,0003	,0007	,0013
30		,0001	,0002	,0004	,0009

Tabla III (continuación)

x \ k	11	12	13	14	15
2	0,9985	0,9994	0,9998	0,9999	1,0000
3	,9907	,9955	,9979	,9991	0,9996
4	,9699	,9834	,9912	,9955	,9977
5	,9312	,9580	,9752	,9858	,9921
6	,8734	,9161	,9462	,9665	,9798
7	,7991	,8576	,9022	,9347	,9577
8	,7133	,7852	,8436	,8893	,9238
9	,6219	,7029	,7729	,8311	,8775
10	,5304	,6160	,6939	,7622	,8197
12	,3636	,4457	,5276	,6063	,6790
14	,2330	,3007	,3738	,4497	,5255
16	,1411	,1912	,2491	,3134	,3821
18	,0816	,1157	,1575	,2068	,2627
20	,0453	,0671	,0952	,1301	,1719
21	,0334	,0504	,0729	,1016	,1368
22	,0244	,0375	,0554	,0786	,1078
23	,0177	,0277	,0417	,0603	,0841
24	,0127	,0203	,0311	,0458	,0651
25	,0091	,0148	,0231	,0346	,0499
26	,0065	,0107	,0170	,0259	,0380
27	,0046	,0077	,0124	,0193	,0287
28	,0032	,0055	,0091	,0142	,0216
29	,0023	,0039	,0066	,0105	,0161
30	,0016	,0028	,0047	,0076	,0119
31	,0011	,0020	,0034	,0055	,0088
32	,0008	,0014	,0024	,0040	,0064
33	,0005	,0010	,0017	,0029	,0047
34	,0004	,0007	,0012	,0021	,0034
35	,0003	,0005	,0009	,0015	,0025
36	,0002	,0003	,0006	,0010	,0018
37	,0001	,0002	,0004	,0007	,0013
38	,0001	,0002	,0003	,0005	,0009
39	,0001	,0001	,0002	,0004	,0006
40		,0001	,0001	,0003	,0005

Tabla III (continuación)

x \ k	16	17	18	19	20	$A_{20}(x)$
4	0,9989	0,9995	0,9998	0,9999	1,000	0,9997
5	,9958	,9978	,9989	,9994	,0,9997	,9990
6	,9881	,9932	,9962	,9979	,9989	,9973
7	,9733	,9836	,9901	,9942	,9967	,9938
8	,9489	,9666	,9786	,9867	,9919	,9876
9	,9134	,9403	,9597	,9735	,9829	,9774
10	,8666	,9036	,9319	,9530	,9682	,9619
12	,7440	,8001	,8472	,8856	,9161	,9109
14	,5987	,6671	,7291	,7837	,8305	,8298
16	,4530	,5238	,5926	,6573	,7166	,7218
18	,3239	,3888	,4557	,5224	,5874	,5968
20	,2202	,2742	,3328	,3946	,4579	,4683
22	,1432	,1847	,2320	,2843	,3405	,3489
24	,0895	,1194	,1550	,1962	,2424	,2472
26	,0540	,0745	,0998	,1302	,1658	,1670
28	,0316	,0449	,0621	,0834	,1094	,1078
30	,180	,0264	,0375	,0518	,0699	,0667
31	,0135	,0200	,0288	,0404	,0552	,0517
32	,0100	,0151	,0220	,0313	,0433	,0396
33	,0074	,0113	,0167	,0240	,0337	,0301
34	,0054	,0084	,0126	,0184	,0261	,0227
35	,0040	,0062	,0095	,0140	,0201	,0169
36	,0029	,0046	,0071	,0106	,0154	,0125
37	,0021	,0034	,0052	,0080	,0117	,0092
38	,0015	,0025	,0039	,0059	,0089	,0067
39	,0011	,0018	,0029	,0044	,0067	,0048
40	,0008	,0013	,0021	,0033	,0050	,0035
41	,0006	,0009	,0015	,0024	,0037	,0025
42	,0004	,0007	,0011	,0018	,0028	,0017
43	,0003	,0005	,0008	,0013	,0020	,0012
44	,0002	,0003	,0006	,0009	,0015	,0010
45	,0001	,0002	,0004	,0007	,0011	,0006

Tabla IV. Distribución de Student T_k

En la tabla se dan los valores de

$$T_k(x) = T_k((x, \infty)) = \frac{\Gamma(k+1)/2}{\sqrt{k\pi}\Gamma(k/2)} \int_x^\infty (1 + t^2/k)^{-(k+1)/2} dt$$

cuando $1 \leq x \leq 20$. Para mayores valores de k se puede utilizar la aproximación (véase el § 2.2, tabla I)

$$T_k(x) \approx \bar{\Phi}(x) = \Phi_{0,1}((x, \infty)). \qquad (2)$$

La exactitud de aproximación (2) cuando $k = 20$ se puede apreciar comparando la última columna de la tabla con la tabla I.

Tabla IV

x \ k	1	2	3	4	5
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3524	,3333	,3257	,3217	,3191
1,0	,2500	,2113	,1955	,1869	,1816
2	,2211	,1765	,1581	,1482	,1419
4	,1974	,1482	,1280	,1170	,1102
6	,1778	,1253	,1039	,0924	,0852
8	,1614	,1068	,0848	,0731	,0659
2,0	,1476	,0917	,0697	,0581	,0510
2	,1358	,0794	,0576	,0463	,0395
4	,1257	,0692	,0479	,0372	,0308
6	,1169	,0679	,0402	,0300	,0241
8	,1092	,0537	,0339	,0244	,0190
3,0	,1024	,0477	,0282	,0200	,0150
2	,0964	,0427	,0247	,0165	,0120
4	,0910	,0383	,0212	,0136	,0096
6	,0862	,0346	,0184	,0114	,0078
8	,0819	,0314	,0160	,0095	,0063
4,0	,0780	,0286	,0140	,0081	,0052
2	,0744	,0261	,0123	,0068	,0045
4	,0711	,0240	,0109	,0058	,0035
6	,0681	,0221	,0097	,0050	,0029
8	,0654	,0204	,0086	,0043	,0024
5,0	,0628	,0199	,0077	,0037	,0020
2	,0605	,0175	,0069	,0033	,0017
4	,0583	,0163	,0062	,0028	,0015
6	,0562	,0152	,0056	,0025	,0012

Tabla IV (continuación)

x \ k	1	2	3	4	5
8	,0543	,0142	,0051	,0022	,0011
6,0	,0526	,0133	,0046	,0019	,0009
2	,0509	,0125	,0042	,0017	,0008
4	,0493	,0118	,0039	,0015	,0007
6	,0479	,0111	,0035	,0014	,0006
8	,0465	,0105	,0033	,0012	,0005
7,0	,0452	,0099	,0030	,0011	,0005
2	,0439	,0094	,0028	,0010	,0004
4	,0428	,0089	,0025	,0009	,0004
6	,0416	,0086	,0024	,0008	,0003
8	,0406	,0080	,0022	,0007	,0003
8,0	,0396	,0076	,0020	,0007	,0002

Tabla IV (continuación)

x \ k	6	7	8	9	10
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3174	,3162	,3153	,3145	,3139
1,0	,1780	,1753	,1733	,1717	,1704
2	,1377	,1346	,1322	,1304	,1280
4	,1055	,1021	,0995	,0975	,0959
6	,0804	,0768	,0741	,0720	,0703
8	,0610	,0574	,0548	,0527	,0510
2,0	,0462	,0428	,0403	,0383	,0367
2	,0350	,0319	,0295	,0277	,0262
4	,0266	,0237	,0216	,0199	,0186
6	,0203	,0177	,0158	,0144	,0132
8	,0156	,0132	,0116	,0104	,0094
3,0	,0120	,0100	,0085	,0075	,0067
2	,0093	,0075	,0063	,0054	,0047
4	,0072	,0057	,0047	,0039	,0034
6	,0057	,0044	,0035	,0029	,0024
8	,0045	,0034	,0026	,0022	,0017
4,0	,0035	,0026	,0020	,0015	,0013
2	,0028	,0020	,0015	,0012	,0009
4	,0023	,0016	,0011	,0009	,0007
6	,0018	,0012	0,009	,0006	,0005
8	,0015	,0010	,0007	,0005	,0004
5,0	,0012	,0008	,0005	,0004	,0003

Tabla IV (continuación)

x \ k	11	12	13	14	15
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3135	,3131	,3127	,3124	,3112
1,0	,1694	,1685	,1678	,1671	,1666
2	,1277	,1266	,1258	,1250	,1244
4	,0945	,0934	,0925	,0916	,0909
6	,0689	,0678	,0668	,0660	,0652
8	,0496	,0485	,0475	,0467	,0460
2,0	,0354	,0343	,0334	,0326	,0320
2	,0250	,0241	,0232	,0225	,0219
4	,0176	,0168	,0160	,0154	,0149
6	,0123	,0116	,0110	,0105	,0100
8	,0086	,0080	,0075	,0071	,0067
3,0	,0060	,0055	,0051	,0048	,0045
2	,0042	,0038	,0035	,0032	,0030
4	,0030	,0026	,0024	,0022	,0020
6	,0021	,0018	,0016	,0014	,0013
8	,0015	,0013	,0011	,0010	,0009
4,0	,0010	,0009	,0008	,0007	,0006

Tabla IV (continuación)

x \ k	16	17	18	19	20
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3119	,3117	,3116	,3114	,3113
1,0	,1661	,1657	,1653	,1649	,1646
2	,1238	,1233	,1228	,1224	,1221
4	,0903	,0898	,0893	,0888	,0884
6	,0646	,0640	,0635	,0630	,0626
8	,0454	,0448	,0443	,0439	,0435
2,0	,0314	,0309	,0304	,0300	,0296
2	,0214	,0210	,0205	,0202	,0199
4	,0145	,0141	,0137	,0134	,0131
6	,0097	,0093	,0090	,0082	,0086
8	,0064	,0061	,0059	,0057	,0055
3,0	,0042	,0040	,0038	,0037	,0035
2	,0028	,0026	,0025	,0024	,0022
4	,0018	,0017	,0016	,0015	,0014
6	,0012	,0011	,0010	,0009	,0009
8	,0008	,0007	,0007	,0006	,0006
4,0	,0005	,0005	,0004	,0004	,0003

Observaciones bibliográficas

Más abajo se aducen algunos comentarios bibliográficos en los que se hacen intentos de seguir la historia de aparición de las ideas y los resultados fundamentales expuestos en este libro. Dichos comentarios no pretenden ser completos y a menudo contendrán referencias no a artículos originales poco abordables, sino a manuales, monografías o artículos de resumen, en los que es más fácil hallar los resultados necesarios. Por ejemplo, en [95] y [57] se ofrecen indicaciones bibliográficas e informaciones históricas más amplias.

Algunos conceptos fundamentales de la estadística matemática surgieron ya a principios del siglo pasado y están relacionados con los nombres de Laplace y Gauss. A finales del siglo pasado, los trabajos de K. Pearson dieron comienzo a un período de desarrollo intenso de dicha ciencia. El mismo ha sido condicionado por las obras fundamentales de R. Fisher, J. Neyman, A. N. Kolmogórov y A. Wald. En la Unión Soviética, el desarrollo de la estadística matemática se halla relacionado, antes que nada, con los nombres de A. N. Kolmogórov y N. V. Smirnov.

Capítulo 1

§§ 2—4. El teorema de Glivenko — Cantelli fue establecido en el año (a Glivenko le pertenece su demostración para una distribución continua, y a Cantelli, para el caso general).

La demostración del teorema 1.2.2 se asemeja a la expuesta en [61], p. 28, y es un caso particular de utilización de un enfoque más general basado en la "aproximación finita" de la clase de conjuntos sujetos a estudio. En su forma completa, este enfoque se ofrece en el Suplemento I, donde ha sido demostrado el teorema 1.4.2. Un enfoque análogo fue examinado independientemente en [27]. La ley del logaritmo reiterado (teorema 1.4.3) fue establecida en [52].

§ 6. Los teoremas 1.6.1 y 1.6.2 de la distribución de $nF_n^*(t)$ se dan en el libro de Feller [32], t. 2, § 3, cap. III. El teorema 1.6.3 de la convergencia del proceso $\sqrt{n}(F_n^*(t) - F(t))$ hacia el puente browniano, demostrado en el Suplemento II, fue establecido por Donsker en [28]. Una demostración algo diferente (en comparación con el Suplemento II) del teorema 1.6.3 se ofrece en la obra de Billingsley [5].

§ 7. La afirmación del ejemplo 1.7.3 acerca de la distribución límite de la estadística $\chi^2(X)$ (ji-cuadrado) fue por primera vez obtenida por K. Pearson (véase [25], p. 454).

§ 8. La afirmación del corolario 1.8.2 constituye el contenido del teorema de Kolmogórov, y la del corolario 1.8.3, el del teorema de Smirnov. Este último también comprende la forma

explícita de la distribución de $\int_0^1 [w^0(t)]^2 dt$, que omitimos debido a su complejidad (véase [78]).

§ 10. Las estimaciones de la densidad que se examinan en este párrafo fueron introducidas por Parzen [72] y Rosenblatt [79]. La bibliografía y el análisis de los resultados en esta dirección se exponen en el trabajo de resumen de Rosenblatt [80] y en el § 25 del libro de Chentsov [19].

Capítulo 2

§ 2. Algunas otras familias paramétricas se describen en el libro de Wilks [93]. Una investigación muy completa de las distribuciones de los términos de la serie variacional fue llevada a efecto B. V. Gnedenko. Una exposición completa de los resultados y una amplia bibliografía al respecto se pueden hallar en la obra de David [26].

§ 4. El método de momentos es, históricamente, el primer método regular de construcción de las estimaciones. El mismo fue propuesto por K. Pearson en 1894.

§ 5. El método del mínimo χ^2 fue propuesto por R. Fisher en 1922.

§ 6. El método de verosimilitud máxima en casos particulares fue empleado aún por Gauss. Como método general para obtener las estimaciones, el mismo fue propuesto por Fisher en 1912 en un artículo breve. Más tarde, en 1925, Fisher estudió las propiedades asintóticas de la e.v.m. en su obra clásica [35].

§§ 7 y 8. Los enfoques expuestos, dedicados a la comparación de las estimaciones, son universalmente reconocidos. Hemos adoptado la demostración del lema 2.7.3 dada en [25]. El concepto de estimación eficiente fue introducido en 1922 por Fisher en [34].

§§ 9 y 10. El concepto fundamental de esperanza matemática condicional fue introducido en 1933 por A. N. Kolmogórov en su obra clásica [54]. Las propiedades de las distribuciones condicionales fueron detalladamente estudiadas en [38], [30] y [84].

§ 11. El enfoque bayesiano ha sido ampliamente utilizado por Laplace aún en el siglo pasado. Este enfoque fue criticado por Fisher, y en los años 20 y 30 de nuestro siglo, el centro de gravedad de las investigaciones se desplazó hacia las estimaciones eficientes y asintóticamente eficientes. Más tarde, a medida que se concebía el papel fundamental del enfoque bayesiano, otra vez comenzó a crecer el interés por este último.

El concepto de estimación minimax se introdujo en la estadística matemática junto con el enfoque de la teoría de los juegos, desarrollado en los trabajos de Borel (1921) y J. Neyman (1928); los teoremas 2.11.1—2.11.3 fueron obtenidos por Hodges y Lehman [44].

§ 12. El concepto fundamental de la estadística suficiente fue introducido en 1922 por R. Fisher en [34], quien, y más tarde J. Neyman [66], propusieron un criterio simple que revela la existencia y el tipo de estadística suficiente. Este criterio lleva el nombre de teorema de factorización de Neyman — Fisher y está representado en el teorema 2.12.1. La estricta demostración del teorema de Neyman — Fisher, desde el punto de vista de la teoría de los conjuntos, fue obtenida tan sólo en 1949 por Halmos y Savage [43].

§ 13. El concepto de σ-álgebra suficiente es más amplio que el concepto de estadística suficiente. Las condiciones necesarias y suficientes para su coincidencia se dan en [95]. Tanto la construcción de las particiones suficientes como el teorema 2.13.1 están relacionados con el trabajo de Lehmann y Scheffe [59] dedicado a la aclaración de las condiciones de existencia y a la construcción de las estadísticas mínimas suficientes. La exposición breve de este artículo se ofrece en [95]. La demostración del teorema 2.13.2 le pertenece a I. S. Borísov.

§ 14. El teorema 2.14.1 fue independientemente obtenido por Blackwell [6] (1947), Rao [75] (1945), [76] (1949) y Kolmogórov [53] (1950). Los autores del teorema 2.14.3 son Rao [76] (1949) y Blackwell [6] (1947).

§ 15. La familia exponencial ha sido mencionada por Fisher aún en [34], pero su importancia teórica fue concebida en los años 30 en las obras de Pitman, Kupman y Darmois. Por eso dicha familia a veces lleva los nombres de estos científicos. El teorema 2.15.2 fue demostrado por Lehmann ([57], p. 183).

§§ 16 y 17. La desigualdad de Rao — Cramer a veces también se denomina desigualdad de información. De hecho, ésta pertenece a Fisher [35], aunque en la forma expuesta fue independientemente obtenida por Frechet [37] en 1943, Rao [74] en 1945 y Cramer [24] en 1946.

Las condiciones de regularidad, necesarias para el cumplimiento de la desigualdad, en los manuales de estadística matemática no siempre se interpretan correctamente. Se trata de las condiciones que aseguran la validez de la derivación respecto al parámetro bajo el signo integral. La demostración de dicha validez a menudo contiene lagunas (véase por ejemplo, [95]) o su exposición no se ofrece en absoluto (por ejemplo, en [86]). En una serie de casos, la misma se menciona en forma de condición [86]), lo cual no es cómodo para la verificación en problemas reales.

Las condiciones de regularidad adoptadas en el libro son muy simples, aunque, por lo visto, no son las más generales (compárense con ([48]). El hecho de que en estas condiciones se pueda derivar bajo el signo integral, fue demostrado en el Suplemento VI escrito a base de los resultados obtenidos por A. I. Sajanenko.

En [95] y [19] se ofrecen distintas generalizaciones de la desigualdad de Rao — Cramer. El concepto de información (de Fisher) fue introducido en [35]. Al demostrar los teoremas 2.16.1A y 2.17.1 nos hemos guiado por los libros [95] y [48].

§§ 18 y 19. A Hotelling y Pitman les pertenece la idea de utilizar las consideraciones invariantes. S. Stein contribuyó considerablemente al desarrollo de la teoría. El contenido principal del teorema 2.18.1 le pertenece a Pitman. Al demostrarlo hemos utilizado las exposiciones en [95] y [48]. El carácter minimax de la estimación de Pitman fue establecido por Girchik y Savage.

§ 20. Los resultados de este párrafo fueron obtenidos por el autor junto con A. I. Sajanenko [13]. Cuando las limitaciones son más rígidas, algunas desigualdades también se pueden obtener de las obras [40] y [18].

§ 21. En el caso paramétrico, la distancia de Kullback — Leibler también se llama función de información de Kullback — Leibler. Al describir las probabilidades de las grandes divergencias de la distribución empírica, I. N. Sanov llegó independientemente a la referida distancia. La idea del amplio uso de la distancia de Hellinger para estudiar las propiedades de la relación de verosimilitud fue adoptada del libro de Ibraguímov y Jasminski [48]. Las demostraciones de los principales teoremas del § 23 también se basan en los resultados de este libro. La demostración del teorema 2.21.3 ha sido considerablemente simplificada por A. I. Sajanenko.

§ 22. El teorema 2.22.1 fue establecido en 1952 por Chapman y Robbins en [17] y en 1952 por Kiefer en [51].

§§ 23—25. Se expone el material de nuestras conferencias, perfeccionado considerablemente después de la aparición del libro de Ibraguímov y Jasminski [48]. Los prefeccionamientos principales están relacionados con la utilización sistemática de la distancia de Hellinger para estimar $\mathbf{M}_\theta Z^{1/2}(u)$. A. I. Sajanenko propuso utilizar $\int \mathbf{M}_\theta |(Z^{3/4}(u))| du$ para estimar $\sup Z(u)$ (véanse los teoremas 2.23.1 y 2.23.2). Aún Fisher, en [35], estableció la normalidad asintótica y la eficacia asintótica de la e.v.m. Condiciones muy generales de la normalidad asintótica de la e.v.m. fueron obtenidas en [48].

La normalidad asintótica de la densidad a posteriori (o de la relación de verosimilitud) fue descubierta por S. N. Bernshtein en 1927. El teorema 2.25.4 pertenece a Bahadur [1]. Los caracteres asintóticamente bayesiano y asintóticamente minimax de la e.v.m. se obtienen fácilmente merced a los resultados del § 2.20. Antes, el carácter asintóticamente bayesiano de la e.v.m. se establecía con limitaciones más rígidas para la densidad de la distribución a priori.

Hemos utilizado, para demostrar los teoremas 2.24.1 y 2.24.2, algunos perfeccionamientos propuestos por A. I. Sajanenko.

§ 26. Se expone una de las variantes del método numérico de Ruffson para determinar el extremo de la función. Véase la exposición con más detalles en [95]. Hemos adoptado el ejemplo 3 del libro de Rao [76].

§ 27. La investigación de la conciliabilidad de la e.v.m. fue comenzada en los años 30 y 40 en los trabajos de Doob [29], Wald [88], Wolfowitz [94] y Cramer [25]. Las principales condiciones de conciliabilidad comprenden, en [88] (además de las condiciones (A_a), (A_c), (A_0)), la pertenencia de $f_t(x)$ a la clase D_0 y la integrabilidad de

$$\int \ln f_t^{\Delta}(x) f_0(x) \mu(dx).$$

En la monografía [48] fueron obtenidas las condiciones de conciliabilidad que utilizan la convergencia

$$\int \sup_{|u| \leq \Delta} (\sqrt{f_{t+u}(x)} - \sqrt{f_t(x)})^2 \mu(dx) \to 0 \quad \text{para} \quad \Delta \to 0.$$

Los resultados de los teoremas 27.1 y 27.2 y de sus corolarios son más generales. El método de demostración es semejante a [88]. La suficiencia de las condiciones (A_0°) y (2.27.2) fue revelada por A. I. Sajanenko.

§§ 28 y 29. Véanse los comentarios a los §§ 23—27. Hemos adoptado el ejemplo 2.28.1. del libro de Van der Waerden [86].

En la exposición de los párrafos 28 y 29 hemos introducido varios perfeccionamientos en comparación con la variante inicial, o sea, mejoras propuestas por A. I. Sajanenko (en particularidad, hemos añadido el teorema 2.29.5). Estas modificaciones permitieron simplificar el texto en los §§ 13—15 del capítulo 3.

§ 30. La estimación sucesiva se expone con más detalles, por ejemplo, en [95].

§§ 31 y 32. Por lo visto, fue Laplace quien introdujo por primera vez los intervalos confidenciales. Aún en 1812 él mostró que se podía invertir respecto a p la afirmación acerca del grado de divergencia de la frecuencia observada y de la probabilidad binomial p, con el fin de hallar el intervalo para los posibles valores de p. En 1927, Wilson dio la justa interpretación de los intervalos confidenciales (la cual no supone la casualidad del parámetro).

En 1930, Fisher, en [36], propuso un método general de determinación de los intervalos confidenciales exactos. En 1937 y 1938 Neyman desarrolló la teoría general de afirmaciones confidenciales y estableció su relación con la teoría de verificación de las hipótesis. La moderna exposición, muy completa, de esta cuestión se puede hallar en el libro de Lehmann [57]. Hemos utilizado esta exposición en el § 3.7.

El teorema 2.32.1 y el lema 2.23.2 le pertenecen a Fisher.

Capítulo 3

Las primeras aplicaciones de los criterios estadísticos remontan a Laplace (final del siglo 18). El uso sistemático de los criterios para verificar las hipótesis se inicia a partir de los trabajos de K. Pearson, quien propuso, en 1900, el criterio χ^2. Los principales conceptos de errores de primero y segundo género fueron introducidos en 1928 por Neyman y Pearson en [68]. Estos mismos autores fueron los primeros en concebir la importancia de las alternativas para elegir racionalmente el criterio. En la obra conclusiva de Neyman y Pearson [69] se desarrolla la teoría del c. u.m.p.

El libro de Lehmann [57] contiene la exposición sistemática de la teoría de verificación de las hipótesis.

§§ 1—3. El lema fundamental de Neyman — Pearson fue obtenido en [69]. Los teoremas 3.1.1 se pueden extraer del libro de Blackwell Girshik [7]. El libro de Lehmann [57] contiene

el teorema 3.2.1. El teorema 3.3.1 de las grandes divergencias le pertenece a Cramer (véase [11]). La estimación de la calidad de los criterios, relacionada con las probabilidades de las grandes divergencias, constituyó la base del concepto de eficacia del criterio de Bahadur. En [3] se exponen los resultados de las investigaciones con arreglo a esta tendencia.

La importancia de la estadística de aportación eficiente fue revelada aún en 1925 en la obra de Fisher [35]. En lo sucesivo, el enfoque relacionado con el estudio de las hipótesis semejantes fue desarrollado intensamente en los trabajos de Le Cam, Roussas y Chíbisov (véanse también los comentarios a los §§ 3.14 y 3.15).

§ 4. La referida concepción general de los criterios estadísticos ha sido universalmente reconocida (véanse [25] y [57]). El concepto de c.u.m.p. fue introducido por Neyman y Pearson en [69]. Aún en el siglo 19, Laplace utilizó el enfoque bayesiano.

§§ 5—8. Los resultados principales de estos párrafos se han tomado del libro de Lehmann [57]. La exposición también es semejante a la de este libro y se distingue por el hecho de que se basa no en el lema generalizado de Neyman — Pearson (lema 3.5.2, véase también [57]), sino en el enfoque bayesiano. Esto simplifica la exposición y la hace más armoniosa.

Ciertas observaciones referentes a los conjuntos confidenciales se exponen en los comentarios a los §§ 2.31 y 3.32.

En el libro de Grenander [39] se examina la posibilidad de extender los resultados principales a los procesos aleatorios.

§ 9. Los autores del teorema 3.9.1 son Hodges y Lehmann [44].

§ 10. El papel fundamental de la relación de verosimilitud en la estadística matemática fue aclarado en los trabajos de Neyman y Pearson [68], [69]. Al estudio del c.r.v. se han dedicado muchos libros. Ciertas tentativas de establecer unas u otras propiedades de optimización asintótica de este criterio se ofrecen en los trabajos [2], [88], [71], [93] y [45].

§ 11. Wald [89] fue quien más contribuyó al desarrollo de la teoría del análisis secuencial. La exposición más completa de los resultados principales, por la cual nos guíamos en nuestro libro, se ofrece en [57].

§ 12. Los criterios de Kolmogórov y ω^2 se exponen en el § 1.8 y en los comentarios a este último. A su vez, algunas modificaciones del criterio de Kolmogórov, que proporcionan la potencia máxima posible, se dan en [16]. El criterio de Moran fue propuesto en [64]. Su potencia para las alternativas semejantes se estudió en [91] y [20].

§ 13. El carácter asintóticamente bayesiano del c.r.v. ue determinado en el trabajo del autor de [10]. Los resultados de la distribución límite de la relación de verosimilitud para la hipótesis principal fueron obtenidos por Wilks [92] y Wald [87] (véase también el libro de Wilks [93]). Wald utilizó la idea de sustituir la hipótesis compleja por una hipótesis promediada. En el trabajo [60] se examina la forma asintótica de los criterios bayesianos. Véanse también los comentarios a los §§ 28 y 29 del capítulo 2.

§§ 14 y 15. Las principales ideas relacionadas con la determinación de los tests asintóticamente óptimos para hipótesis semejantes se exponen en las obras de Wald [87], Le Cam, Roussas (véase el libro de Roussas [81]) y Chíbisov [22]. En el libro [14] se analiza la posibilidad de extender los resultados principales al caso del parámetro de dimensión infinita (es decir, a los procesos aleatorios). La forma de exposición de los §§ 14 y 15 está poco relacionada con los trabajos citados. En el libro [87] de Wald se ofrece la reducción del problema inicial A a un problema B para el parámetro de distribución normal al determinar los criterios óptimos de los principales tipos de problemas examinados en el § 14. La afirmación del teorema 3.15.4 acerca de la distribución de la estadística $2 \ln R_1(X)$ para la hipótesis H_1 se examina en [93]. Véanse también los comentarios a los §§ 28 y 29 del capítulo 2.

§§ 16 y 17. En el año 1900, K. Pearson propuso el criterio χ^2, al cual se han dedicado muchos libros (véase, por ejemplo, la monografía especial de Lancaster [56]). El examen de las diversas propiedades de la optimización se expone en [87], [71], [93], [45], etc. El comportamiento de la potencia del criterio χ^2 al aumentar el número de grupos se analiza, por ejemplo,

en [12] y [21]. Los ejemplos 3.16.1 y 3.17.2 se han adoptado del libro de Cramer [25], y el ejemplo 3.17.1, del libro de Rao [76].

§ 18. Al estudiar la estabilidad de las decisiones estadísticas es muy difícil seguir la etapa inicial de ese estudio. Las investigaciones posteriores se basan en los trabajos de Takeuchi, Hodges y Lehmann. En el libro [47] de Huber se hace un resumen detallado de dicha tendencia.

Capítulo 4

§ 1. El criterio χ^2 en el problema del ejemplo 4.1.1, el criterio de Student en el problema del ejemplo 4.1.3 y el criterio de Fisher en los problemas de los ejemplos 4.1.4 y 4.1.5 se utilizan muy a menudo. En el libro [57] de Lehmann se dan otras propiedades de optimización de estos criterios. El ejemplo 4.1.1A se ha tomado del libro [76]. Hay muchos libros (véase [57]) dedicados al problema de Beherns — Fisher (ejemplo 4.1.6).

§ 2. Gnedenko y Koroliuk (véase [32]) hallaron la distribución exacta de la estadística $D_{n,n}$, y Smirnov, la distribución límite de la estadística D_{n_1,n_2}. El teorema 4.2.2. fue demostrado por primera vez en [62] con ayuda del método de momentos. Los criterios de signos y de Wilkoxon también se ofrecen en [41].

§§ 3 y 4. Los problemas de regresión y análisis de varianza se exponen más detalladamente en las monografías especiales de Seber [83] y Scheffe [82]. Véanse asimismo [25], [57] y [76].

§ 5. La observación acerca de la optimización asintótica del criterio (4.5.3) fue tomada de [10].

Capítulo 5

En matemática, la tendencia relacionada con la teoría de los juegos surgió tras la publicación de los trabajos de Borel en 1921 y de von Neumann en 1928. En la estadística matemática, como trabajo inicial, que preparó el uso de la teoría de los juegos, puede considerarse la obra clásica de Neyman y Pearson [70], en la que se enuncian muchas ideas fundamentales de la teoría de las decisiones estadísticas. Wald contribuyó considerablemente al desarrollo de la teoría general de las decisiones estadísticas. En su libro conclusivo [90] se exponen los postulados fundamentales de esta teoría. No obstante, la teoría matemática general de los juegos adquirió su pleno desarrollo en el libro de von Neumann y Morgenstern [65].

Los fundamentos de la teoría de los juegos estadísticos plantean de una forma muy accesible en los libros de Girshik y Blackwell [7] y de Ferguson [33].

§ 2. El libro de McKinsey [63] constituye una introducción relativamente completa a la teoría ordinaria de los juegos.

§§ 3 y 4. En [7] y [33] se da una descripción más completa de los fundamentos de la teoría de los juegos estadísticos. En estos libros, dos teoremas fundamentales de la teoría de los juegos estadísticos sólo se demuestran en el caso particular, para los conjuntos discretos D y θ. Ello se explica por el hecho de que la exposición en el caso general es muy compleja (véase [90]). En el Suplemento VIII se da la demostración más simple que conocemos de tales teoremas, la cual fue hallada por A. I. Sajanenko.

El papel del enfoque bayesiano en distintos tiempos se evaluaba de manera diferente. El mismo ha sido ampliamente utilizado por Laplace en el siglo pasado. Después fue criticado por Fisher, y en los años 20 y 30 de nuestro siglo, el centro de gravedad se desplazó hacia las estimaciones eficientes y asintóticamente eficientes. Más tarde, a medida que se concebía la importancia fundamental del referido enfoque, otra vez comenzó a crecer el interés por él. Esa importancia fundamental es aclarada en los teoremas 5.3.1 y 5.3.2.

§ 5. El concepto fundamental de estadística suficiente fue introducido por R. Fisher [34] en el año 1922. R. Fisher [34] y más tarde J. Neyman [66] propusieron un criterio simple

que revela la existencia y el tipo de estadística suficiente. Este criterio es conocido con el nombre de teorema de factorización de Neyman — Fisher y está representado en el teorema 2.12.1. La estricta demostración del teorema de Neyman — Fisher, desde el punto de vista de la teoría de los conjuntos, fue obtenida tan sólo en 1949 por Halmos y Savage [43].

El concepto de σ-álgebra suficiente es más amplio que el concepto de estadística suficiente. Las condiciones necesarias y suficientes para su coincidencia se dan en [95]. El teorema 5.5.1 (primero para la función cuadrática de pérdidas) fue independientemente obtenido por Blackwell *[6] (1947)*, Rao *[74] (1945)*, *[75] (1949)* y *Kolmogórov [53] (1950)*. Las generalizaciones para el caso de función arbitraria de pérdidas están íntimamente ligadas a los nombres de Lehmann y Scheffe [95].

A Hotelling y Pitman les pertenece la idea de utilizar las consideraciones invariantes. Ch. Stein (véanse [95] y [48]) contribuyó considerablemente al desarrollo de la teoría.

En [95] se ofrecen datos más detallados acerca del carácter no desplazado.

§ 6. El libro [48] de Ibraguímov y Jasminski contiene resultados semejantes a los teoremas de este párrafo.

§ 7. El carácter asintóticamente bayesiano del c.r.v. fue establecido en el trabajo del autor de [10]. Los resultados de la distribución límite de la relación de verosimilitud para la hipótesis principal fueron obtenidos por Wilks [92] y Wald [87] (véase también el libro de Wilks [93]). Wald utilizó la idea de sustituir la hipótesis compleja por una hipótesis promediada. El tipo asintótico de criterios bayesianos se expone en [60].

§ 8. Las principales ideas relacionadas con la determinación de los tests asintóticamente óptimos para hipótesis semejantes se examinan en los trabajos de Wald [87], Le Cam, Roussas (véase el libro [81] de Roussas) y Chíbisov [22]. En [15] se estudia la posibilidad de extender los resultados principales al caso de un parámetro de dimensión infinita (es decir, a los procesos aleatorios). La forma de exposición del § 8 y de los §§ 14 y 15 del capítulo 3 está poco relacionada con los trabajos citados. La reducción del problema inicial A a un problema B (para el parámetro de distribución normal), al determinar los criterios óptimos para los principales tipos de problemas, se analiza en el trabajo de Wald [87].

Suplemento VII

Fue A. A. Mogulski quien propuso utilizar el lema 1 para demostrar el teorema 2.28. La demostración de este lema se remonta a S. L. Sóbolev. La demostración del lema 1 también se puede obtener fácilmente utilizando los resultados de [96]. En la edición rusa del libro se da otra demostración del teorema 2.28, la cual utiliza ciertas ideas de A. N. Kolmogórov acerca de la estimación de la distribución del máximo del proceso aleatorio.

Suplemento VIII

La demostración de dos teoremas fundamentales de la teoría de los juegos estadísticos se ofrece en [90] y, para suposiciones más particulares, en [7] y [33]. En el libro presente se expone el enfoque de la demostración propuesta por A. I. Sajanenko. Su parte central consta de los lemas 2 y 3. De hecho, el lema 2 no está relacionado con el carácter estadístico del juego, se basa en los teoremas de Hahn — Banach y de Riss y por su idea se asemeja a los razonamientos utilizados, por ejemplo, en [31]. La demostración del lema 3 se basa en los teoremas de Kolmogórov [54] y Prójorov [5].

Al trazar las tablas I—IV se utilizó el libro de Bolshev y Smirnov [8].

Bibliografía

1. *Bahadur R. R.* On Fisher's bound for asymptotic variances. Ann. Math. Statist., 1964, 35, 4, 1545—1552.
2. *Bahadur R. R.* An optimal property of the likelihood ratio statistic, Proc. 5-th Berkeley Sympos. Math. Statist. Prob. — Berkeley — Los Angeles, v. 1, 1965, 27—40.
3. *Bahadur R. R.* Some limit theorems in statistics. — Philadelphia: S.I.A.M., 1971.
4. *Bahadur R. R., Lehmann E. L.* Two comments on "Sufficiency and statistical decision functions". — AMS. 1955, 26, 139—141.
5. *Billingsley P.* Convergence of probability measures, N.Y., Wiley, 1968.
6. *Blackwell D.* Conditional expectation and unbiased sequential estimation. — Ann. Math. Statist., 1947, 18, 105—110.
7. *Blackwell D., Girshik M. A.* Theory of games and statistical decisions, N.Y., Wiley, 1954.
8. *Большев Л. Н., Смирнов Н. В.* Таблицы математической статистики. — М.: Наука, 1965.
   (*Bólshev L. N., Smirnov N. V.* Tablas de estadística matemática.)
9. *Боровков А. А.* Вероятностные процессы в теории массового обслуживания. — М.: Наука, 1972.
   (*Borovkov A. A.* Procesos probabilísticos en la teoría de las colas.)
10. *Боровков А. А.* Асимптотически оптимальные тесты для проверки сложных гипотез. — Теория вероятн. и ее примен., 1975, 20, 3, 463—487.
    (*Borovkov A. A.* Tests asintóticamente óptimos para verificar las hipótesis compuestas. — Teoría de las probabilidades y su aplicación.)
11. *Боровков А. А.* Теория вероятностей. — М.: Наука, 1976,
    (*Borovkov A. A.* Teoría de las probabilidades.)
12. *Боровков А. А.* О мощности критерия при увеличении числа групп. — Теория вероятн. и ее примен., 1977, 22,2, 375—379.
    (Acerca de la potencia del criterio al aumentar el número de grupos. Teoría de las probabilidades y su aplicación.)
13. *Боровков А. А., Саханенко А. И.* Неравенства типа Рао-Крамера для байесовского риска. — Теория вероятн. и ее примен. 1980, 25, 1, 207—209.
    (*Borovkov A. A., Sajanenko A. I.* Desigualdades del tipo de Rao — Cramer para el riesgo bayesiano.)
14. *Боровков А. А., Саханенко А. И.* Об асимптотически оптимальных тестах для проверки сложных гипотез. — Труды Института математики СО АН СССР, 1981, т. 1.
    (Acerca de los tests asintóticamente óptimos para verificar las hipótesis compuestas.)

15. *Боровков А. А., Саханенко А. И.* Об асимптотически оптимальных тестах для проверки сложных близких гипотез. — Труды Института математики СО АН СССР, 1982, т. 1.
(*Borovkov A. A., Sajanenko A. I.* Acerca de los tests asintóticamente óptimos para verificar las hipótesis compuestas semejantes.)
16. *Боровков А. А., Сычева Н. М.* О некоторых асимптотически оптимальных непараметрических критериях. — Теория вероятн. и ее примен., 1968, 13, 3, 385—418.
(*Borovkov A. A., Sicheva N. M.* Acerca de algunos criterios no paramétricos asintóticamente óptimos.)
17. *Chapman D. G., Robbins H. E.* Minimum variance estimation without regularity assumptions. — Ann. Math. Statist., 1951, 22, 581—586.
18. *Ченцов Н. Н.* Об оценке неизвестного многомерного нормального распределения. — Теория вероятн. и ее применен., 1967, 12, 4, 619—633.
(*Chentsov N. N.* Acerca de la estimación de la distribución normal multidimensional media desconocida. Teoría de las probabilidades y su aplicación.)
19. *Ченцов Н. Н.* Статистические решающие правила и оптимальные выводы. — М.; Наука, 1972.
(*Chentsov N. N.* Reglas estadísticas de decisión y deducciones óptimas.)
20. *Чибисов Д. М.* О критериях согласия, основанных на выборочных промежутках. — Теория вероятн. и ее примен., 1961, 6, 1, 354—358.
(Chíbisov D. M. Acerca de los criterios de aceptación basados en los intervalos muestrales. — Teoría de las probabilidades y su aplicación.)
21. *Чибисов Д. М., Гванцеладзе Л.* О критериях согласия, основанных на сгруппированных данных. — В кн.: III Советско-японский симпозиум по теории вероятностей. — Ташкент: Фан, 1975, 183—185.
(*Chíbisov D. M., Gvantseladze L.*. Acerca de los criterios de aceptación basados en los datos agrupados.)
22. *Chíbisov D. M.* Transition to the limiting process for deriving asymptotically optimal tests. Sankhya, 1969, A31, 3, 241—258.
23. *Cox D., Hinkly D.* Theoretical statistics, Chapman and Hall, London, 1974.
24. *Cramer H.* A contribution to the theory of statistical estimation. — Aktuariestidskrift, 1946, 29, 458—463.
25. *Cramer H.* Mathematical Methods of Statistics, 1946.
26. *David H. A.* Order Statistics, N.Y., Wiley, 1976.
27. *De Hardt T.* Generalizations of the Glivenko — Cantelli theorem. — Ann. Math. Stat., 1971, 42, 2050—2055.
28. *Donsker M.* Justifications and extension of Doob's heuristic approach to the Kolmogórov — Smirnov theorems. — Ann. Math. Statist., 1952, 23, 277—281.
29. *Doob J. L.* Probability and statistics. — Trans. Amer. Math. Soc., 1934, 36, 4, 759—775.
30. *Doob J. L.* Stochastic Processes, N.Y., Wiley, 1953.
31. *Edwards R. E.* Functional analysis, HOLT, Rinehart and Winston, New York, Chicago, San Francisco, Toronto, London, 1965.
32. *Feller W.* An Introduction to Probability Theory and its Applications, vols. I, II, N.D., Wiley Eastern, 1972.
33. *Ferguson J. S.* Mathematical statistics. A decision theoretic approach. — New York and London: Academic Press, 1967.
34. *Fisher R. A.* On the mathematical foundations of theoretical statistics. — Phil. Trans. Roy. Soc. A, 1922, 222, 309—368.
35. *Fisher R. A.* Theory of statistical estimation. — Proc. Camp. Phil. Soc., 1925, 22, 700—725.

36. *Fisher R. A.* Inverse probability. — Proc. Cambridge Phill. Soc., 1930, 26, 528—535.
37. *Frechet M.* Rev. Intern. de Stat. 1943, 182.
38. *Гихман И. И., Скороход А. В.* Введение в теорию случайных процессов. — М.: Наука, 1977.
(*Guijmán I. I., Skorojod A. V.* Introducción a la teoría de los procesos aleatorios, en ruso.)
39. *Grenander U.* Stochastic processes and statistical interference, Ark. Math., 1,3, 1960, 195—277.
40. *Гусев С. И.* Асимптотические разложения, связанные с некоторыми статистическими оценками в гладком случае. — Теория вероятн. и ее примен. — 1976, 21, 1, 16—33.
(*Gúsev S. I.* Desarrollos asintóticos relacionados con algunas estimaciones estadísticas en un caso suave. — Teoría de las probabilidades y su aplicación.)
41. *Hajek J., Sidak Z.* Theory of rank tests. Academía Publishing House of the Czechoskovak Academy of Sciences, Prague, 1967.
42. *Halmos P. R.* Measure Theory, Princeton, N.Y., Van Nostrand, 1962.
43. *Halmos P. R., Savage L. J.* Application of the Radon-Nikodym theorem to the theory of sufficient statistics. — Ann. Math. Statist., 1949, 20, 225—241.
44. *Hodges J., Lehmann E.* Some problems in minimax estimation. — Ann. Math. Statist., 1950, 21, 2, 182—197.
45. *Hoeffding W.* Asymptotically optimal test for multinomial distributions. — Ann. Math. Statist., 1965, 36, 2, 369—401.
46. *Hotelling H.* The generalization of student's ratio. Ann. Math. Statist., 1931, 2, 360—378.
47. *Huber P. J.* Robust statistics: a review. — Ann. Math. Statist., 1972, 43, 1041—1067.
48. *Ибрагимов И. А., Хасьминский Р.* Асимптотическая теория оценивания. — М.: Наука, 1979.
(*Ibraguímov I. A., Jasminski P.* Teoría asintótica de la estimación.)
49. *Kendall H. G.* Stuart A. Interference and Relationship, Charles Griffin & Company Limited, London, 1967.
50. *Kendall M. G., Stuart A.* The advanced theory of statistics, vol. 2, Ch. Griffin & Company Limited, London, 1961.
51. *Kiefer J.* On minimum variance estimators. — Ann. Math. Statist., 1952, 23, 627—629.
52. *Kiefer J.* On large deviations of the identically distributed functions of vector chance variables and LIL. — Pacif. J. Math. 1961, 11, 2, 649—660.
53. *Колмогоров А. Н.* Несмещенные оценки. — Изв. АН СССР, сер. мат., 1950, 303.
(*Kolmogórov A. N.* Estimaciones no desplazadas.)
54. *Колмогоров А. Н.* Основные понятия теории вероятностей — М.: Наука, 1974.
(*Kolmogórov A. N.* Conceptos fundamentales de la teoría de las probabilidades.)
55. *Kullback S., Leibler R. A.* On information and sufficiency. — Ann. Math. Statist., 1951, 22, 79—86.
56. *Lancaster H. O.* The chi-squared distribution. — N.Y., Wiley, 1969.
57. *Lehmann E. L.* Testing statistical hypotheses, John Wiley, New York, 1959.
58. *Lehmann E. L.* Theory of point estimation, Wiley, N.Y., 1983.
59. *Lehmann E. L., Scheffe H.* Completeness, similar regions and unbiased estimatiion, — Pt. I. Sankhya, 1950, 10, 305—340.
60. *Lindley D.* The use of prior probability distributions in statistical interference and decision. — Proc. 4-th Berkeley Sympos. Math. Statist. Prob., Berkeley — Los Angeles, v. 1, 1960, 453—468.
61. *Loeve M.* Probability Theory, 2nd ed., D. Van Nostrand Co., 1960.
62. *Mann H. B., Whitney D. R.* On a test whether one of two random variables is stochastically lerger than the other. — Ann. Math. Statist. 1947, 18, 50.

63. *McKinsey J. C. C.* Introduction to the theory of games. — McGraw-Hill, N.Y., 1952.
64. *Moran P. A. P.* The random division of an interval, — J. Roy. Stat. Soc., Suppl., 1947, 9, 92—98.
65. *Von Neumann.* Morgenstern, Theory of Games and Economic Behaviour, Princeton, Princeton University Press, 1953.
66. *Neyman J.* Su un teorema concernente le cosidette statistiche sufficienti. — Inst. Ital. Atti. Giorn., 1935, 6, 320—334.
67. *Neyman J.* First course in probability and statistics,, Holt, Rinehart and Winston, JNC, N.Y., 1950.
68. *Neyman J., Pearson E. S.* On the use and interpretation of certain test criteria. — Biometrica, 1928, 20A, 175—240, 263—294.
69. *Neyman J., Pearson E. S.* On the problem of the most efficient test of statistical hypotheses, — Phil. Trans. Roy. Soc., Ser. A, 1933, 231, 289—337.
70. *Neyman J., Pearson E. S.* The testing of statistical hypotheses in relation to probabilities a priori. — Proc. Camb. Phil. Soc. 1933, 24, 492—510.
71. *Oosterhoff J., W. R. van Zwet.* The likelihood ratio test for the multinomial distribution. Proc. 6-th Berkeley Sympos. Math. Statist. Prob., Berkeley — Los Angeles, v. 1, 1970, 31—50.
72. *Parzen E.* On estimation of a probability density function and mode. — Ann. Math. Statist., 1962, 33,3, 1065—1076.
73. *Pitman E. J. G.* The estimation of the location and scale parameters of a continuous population of any given form. Biometrica, 1938, 30, 391—421.
74. *Rao C. R.* Information and accuracy attainable in estimation of statistical parameters. Bull. Calcutta Math. Soc., 1945, 37, 81—91.
75. *Rao C. R.* Sufficient statistics and minimum variance estimates. — Proc. Cambr. Phill. Soc., 1949, 45, 213—218.
76. *Rao C. R.* Linear Statistical Interference and its Applications. 2nd ed., Wiley: N.Y., 1973.
77. *Скороход А. В.* Случайные процессы с независимыми приращениями. — М.: Наука, 1964.
(*Skorojod A. V.* Procesos aleatorios con incrementos independientes.)
78. *Смирнов Н. В.* О распределении ω^2-критерия Мизеса. — В кн.: Смирнов Н. В. Теория вероятностей и математическая статистика. Избранные труды. — М.: Наука, 1970.
(*Smirnov N. V.* Acerca de la distribución del ω^2-criterio de Mises. — En el libro: Smirnov N. V. Teoría de las probabilidades y estadística matemática.)
79. *Rosenblatt M.* Remarks on some nonparametric estimates of a density function. — Ann. Math. Statist., 1956, 27, 3, 832—837.
80. *Rosenblatt M.* Curve estimation. — Ann. Math. Statist., 1971, 42, 6, 1815.1842.
81. *Roussas G.* Contiguity of probability measures. Cambridge University Press, 1972.
82. *Scheffe H.* The analysis of variance, Wiley; N.Y., 1959.
83. *Seber G. A.* Linear regression analysis, Wiley; N.Y., 1977.
84. *Ширяев А. Н.* Вероятность. — М.: Наука, 1980.
(*Shiriaev A. N.* Probabilidad, en ruso.)
85. *Sídorov V. A.* y otros. Measurement of the $\varphi \to \pi^+\pi^-$ branching ratio, — Physics Letters, 1981, 99B, 1, 62—65.
86. *van der Waerden.* Mathematische Statistik. Springer-Verlag, 1957.
87. *Wald A.* Tests of statistical hypotheses concerning several parameters when the number of observations is large. — Trans. Amer. Math. Soc., 1943, 54, 3, 426—482.
88. *Wald A.* Note on the consistency of the maximum likelihood estimate. — Ann. Math. Statist., 1949, 20, 595—601.
89. *Wald A.* Sequential analysis, Wiley; N.Y., 1947.

90. *Wald A.* Statistical decision functions, New York, 1950.
91. *Weiss L.* The asymtotic power of certain tests of fit based on sample spacings. — Ann. Math. Statist., 1957, 28, 3, 783—786.
92. *Wilks S. S.* The large sample distribution of the likeli-hood ratio for testing composite hypothesis. — A. Math. Statist., 1938, 9, 60—62.
93. *Wilks S. S.* Mathematical Statistics, Wiley, N.Y., 1962.
94. *Wolfowitz J.* On Wald's proof of the consistency of the maximum likelihood estimate. — Ann. Math. Statist., 1949, 20, 601—602.
95. *Zacks S.* The theory of statistical interference, Wiley, N.Y., 1971.
96. *Фаддеев Д. К., Вулих Б. З.* и др., Избранные главы анализа и высшей алгебры., Ленинград, Изд-во ЛГУ, 1981, 199 с.
(*Faddéev D. K., Vúlij B. Z.* y otros. Capítulos escogidos del análisis y el álgebra superior.)

Designaciones principales

Las designaciones se dan en orden alfabético: primero el alfabeto ruso, después el latino y el griego. Al final se ofrecen los símbolos matemáticos.

(A_0), condición de correspondencia biunívoca entre el conjunto paramétrico Θ y la familia de distribuciones $\mathscr{P} = \{\mathbf{P}_\theta\}_{\theta\in\Theta}(\mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_2}$ si $\theta_1 \neq \theta_2)$

(A_c), condición consistente en que el conjunto paramétrico Θ es compacto

(A_μ), condición en virtud de la cual todas las distribuciones de la familia $\mathscr{P} = \{\mathbf{P}_\theta\}$ son dominadas por la medida μ (existe la densidad $f_\theta = d\mathbf{P}_\theta/d\mu$)

b, $b(\theta)$, desplazamiento

$\mathfrak{B}$, σ-álgebra de los conjuntos de Borel sobre la recta R

$\mathfrak{B}_{\mathscr{X}}$, σ-álgebra en el espacio de fase $\mathscr{X}$ (de los conjuntos de Borel si $\mathscr{X} = R^m$

$\mathbf{B}_p$, distribución polinomial (incluyendo la distribución de Bernoulli)

$C(a, b)$, espacio de las funciones continuas en $[a, b]$.

c.a.b., criterio asintóticamente bayesiano

c.a.u.m.p., criterio asintóticamente uniforme más potente

c.d., casi por doquier

c.m.p., criterio más potente

c.r.v., criterio de la relación de verosimilitud

c.t., casi todos (los)

c.u.m.p., criterio uniformemente más potente

$D(a, b)$, espacio de las funciones en $[a, b]$, continuas a la izquierda (en el punto a a la derecha) y que sólo tienen un número finito de saltos

D, espacio de las estrategias del primer jugador (en el cap. 4)

D_θ, varianza de la distribución $\mathbf{P}_\theta$

$\mathscr{D}$, espacio de las funciones de decisión en un juego estadístico

E, matriz unidad

e.m.c., esperanza matemática condicional

e.v.m., estimación de la verosimilitud máxima

$\mathscr{E}$, familia exponencial de las distribuciones

$f_\theta(x)$, densidad de la distribución $\mathbf{P}_\theta$ respecto a la medida μ

$f_\theta(X)$, función de verosimilitud igual (por definición) a $\prod_{i=1}^{n} f_\theta(x_i)$

$F(x)$, por regla general, la función de distribución correspondiente a la distribución P

$F_n^*(x)$, función empírica de distribución

F_{k_1,k_2}, distribución de Fisher

G, grupo de transformaciones de $\mathscr{X}$ en sí, correspondiente a la familia invariante
h_ε, cuantila de la distribución χ^2
H_i, hipótesis
H_k, distribución χ^2
I_x, distribución concentrada en el punto x
$I(\theta) = \|I_{ij}(\theta)\|$, $I_{ij}(\theta) = \mathbf{M}_\theta \frac{\partial}{\partial \theta_i} l(x_1, \theta) \frac{\partial}{\partial \theta_j} l(x_1, \theta)$, matriz de información de Fisher
I_A, indicador del conjunto A
K_b, clase de estimaciones con desplazamiento $b = b(\theta)$
K_0, clase de estimaciones no desplazadas
$\mathscr{K}_0$, clase de estimaciones asintóticamente no desplazadas
$\mathscr{K}^0$, clase de estimaciones asintóticamente centrales
$K_{0,2}$, clase de estimaciones asintóticamente normales θ^*, para las cuales $\mathbf{M}_\theta n(\theta^* - \theta)^2 \to \sigma^2(\theta)$, donde $\sigma^2(\theta)$ es la varianza de la distribución normal límite para $\sqrt{n}(\theta^* - \theta)$
K_ε, (en el cap. 3) clase de criterios de dimensión ε (de nivel $1 - \varepsilon$)
K_ε, clase de criterios no desplazados de dimensión ε
$\mathscr{K}_\varepsilon$, clase de criterios de nivel asintótico $1 - \varepsilon$
$\mathscr{K}_\varepsilon^{Q_1}$, clase de criterios de dimensión ε para el enfoque parcialmente bayesiano
$\mathscr{K}_\varepsilon^{Q_1}$, clase de criterios de dimensión asintótica ε para el enfoque parcialmente bayesiano
$K_{\alpha_1,\dots,\alpha_{r-1}}$, clase de criterios con valores fijos α_i de las probabilidades de los errores de i-ésimo género, $i = 1, \dots, r - 1$
$K_{\alpha,\sigma}$, distribución de Cauchy
$l(x, \theta) = \ln f(x)$
$L(X, \theta) = \ln f_\theta(X)$, función logarítmica de verosimilitud
L_{α,σ^2}, distribución lognormal
$\mathbf{M}_\theta$, esperanza matemática de la distribución $\mathbf{P}_\theta$
$\mathbf{M}(\xi/\mathfrak{U})$, esperanza matemática condicional ξ respecto a la σ-álgebra $\mathfrak{U}$
$\mathbf{M}(\xi/\eta)$, esperanza matemática condicional ξ respecto a la variable aleatoria η
n, volumen de la muestra
$N_\mathbf{P}$, N_F, portador de la distribución $\mathbf{P}$ con la función de distribución F
$\mathbf{P}$, símbolo de la distribución, utilizado en distintos sentidos
$P(B/y)$, distribución condicional
$\mathbf{P}_n^*$, distribución empírica
$\mathbf{P}_\theta$, distribución dependiente del parámetro
$\mathscr{P}$, familia de distribuciones
$\mathbf{Q}$, estrategia randomizada de la "naturaleza" (distribución a priori de θ)
$\mathbf{Q}_x$, distribución a posteriori de θ
$\overline{\mathbf{Q}}$, la peor distribución de θ (estrategia minimáx de la "naturaleza")
$q(t/X)$, densidad de distribución a posteriori de θ
R, recta real
R^m, espacio euclídeo m-dimensional
(R), condición de regularidad de la familia paramétrica en cuya virtud la función $\sqrt{f_\theta(x)}$ es continuamente derivable respecto a θ, y la información de Fisher es positiva y continua
(RR), condiciones de regularidad de la familia paramétrica, que exigen el cumplimiento de las condiciones (A_0), (A_c) y (R), así como de la derivabilidad continua de segundo orden de la función $l(x, \theta)$ y de la existencia de la mayorante $l(x) \geqslant |l''(x, t)|$, para la cual la integral de $\mathbf{M}_\theta l(x_1)$ converge uniformemente hacia Θ
$S = S(X)$, estadística
S^2, varianza empírica
S_X^2, varianza empírica correspondiente a la muestra X

$S_0^2 = \frac{n}{n-1} S^2$

T_k, distribución de Student

$U_{a,b}$, distribución uniforme en $[a, b]$

$u^* = \sqrt{n}(\hat{\theta}^* - \theta)$, estimación normalizada de verosimilitud máxima

$w(t)$, (no siempre) proceso wieneriano

$w^0(t)$, puente browniano

$w^n(t)$, proceso empírico

x_i — elemento de la muestra

$X = X_n = (x_1, \ldots, x_n)$ — muestra de volumen n

$[X_\infty]_n = X_n$ — parte de una muestra infinita, constituida por primeros h elementos de esta última

$x_{(i)}$, i-ésimo elemento de una serie variacional

$\bar{x}$, media empírica

$\mathscr{X}$, espacio al cual pertenecen observaciones (espacio de fase de la muestra)

$(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, P)$, espacio probabilístico muestral correspondiente a una observación

$(\mathscr{X}^n, \mathfrak{B}_{\mathscr{X}}^n, P)$, espacio probabilístico muestral correspondiente a la muestra de volumen n

$x = (x_1, \ldots, x_n)$, elemento de $\mathscr{X}^n$

$\alpha_i(\pi)$ — probabilidad del error de i-ésimo género del criterio π

$\beta(\delta)$, potencia del criterio δ

$\beta_\pi(\theta)$, función de potencia del criterio π

$\beta_{\lambda_1,\lambda_2}$, distribución beta

$\Gamma_{\alpha,\lambda}$, distribución gamma

$\delta = \delta(X)$, (en el cap. 3) regla (criterio) de decisión o (en cap. 5) función de decisión

δ, estrategia del primer jugador

ζ_p, cuantila de orden p

ζ_p^*, cuantila muestral de orden p

θ, parámetro (estrategia de la "naturaleza")

$\theta^{\pm}$, fronteras del intervalo confidencial para el parámetro θ

θ^*, estimación del parámetro θ

θ_Q^*, estimación bayesiana del parámetro θ, la cual corresponde a la distribución a priori Q

$\bar{\theta}^*$, estimación minimáx del parámetro θ

$\hat{\theta}^*$, estimación de verosimilitud máxima del parámetro θ

Θ, conjunto de valores posibles del parámetro θ

Θ^*, conjunto confidencial

λ_ε, cuantila de la distribución normal

$\pi = \pi(X)$, (en el cap. 3) criterio randomizado o (en los caps. 3 y 5) regla (criterio) randomizada de decisión

π, estrategia randomizada del primer jugador

π_Q, criterio (estrategia) bayesiano correspondiente a la distribución a priori Q

$\pi_{Q_1 Q_2}$, criterio bayesiano para el enfoque parcialmente bayesiano

$\bar{\pi}$, criterio (estrategia) minimáx

$\hat{\pi}$, criterio de la relación de verosimilitud

π^0, criterio uniformemente más potente

Π_λ, distribución de Poisson

Φ_{a,σ^2}, distribución normal

$\Phi(x)$, función de la distribución estándar normal

$\underset{d}{=}$, símbolo que significa la coincidencia de las distribuciones de muestras o de variables aleatorias

$\xrightarrow[P]{}$, signo de convergencia en probabilidad

$\xrightarrow[c.s.]{}$, signo de convergencia casi segura (con probabilidad 1)

$\Rightarrow$, signo de convergencia débil de las distribuciones (se utiliza tanto entre las variables aleatorias como entre las distribuciones)

$\in$, signo utilizado entre las designaciones de la muestra (de la variable aleatoria) y de la distribución: significa que la muestra fue extraída de una distribución dada (la variable aleatoria tiene una distribución dada)

$\Subset$, signo de convergencia débil. La relación $\xi_n \Subset \mathbf{P}$ quiere decir que la distribución ξ_n converge débilmente hacia $\mathbf{P}$ cuando $n \to \infty$

Indice alfabético de materias

A NUESTROS LECTORES:

Mir edita libros soviéticos traducidos al español, inglés, francés, árabe y otros idiomas extranjeros. Entre ellos figuran las mejores obras de las distintas ramas de la ciencia y la técnica; manuales para los centros de enseñanza superior y escuelas tecnológicas, literatura sobre ciencias naturales y médicas. También se incluyen monografías, libros de divulgación científica y ciencia-ficción.

Dirijan sus opiniones a la Editorial Mir, 1 Rizhski per., 2, 129820, Moscú, I-110, GSP, URSS.